

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 4, 2002, 19:02:26 ; Search time 81 Seconds  
(without alignments)  
26.321 Million cell updates/sec

Title: US-09-462-625-2\_COPY\_145\_160  
Perfect score: 86  
Sequence: 1 RGFLRSNYEVKGRDV 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

```

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	86	100.0	181	AAW23723	Murine granulocyte
2	86	100.0	182	AAW00770	Mouse tag7 clone p
3	72	83.7	190	AAW23722	Bovine granulocyte
4	64	74.4	191	AAW00771	Human tag7 clone p
5	64	74.4	196	AAW24022	Human PRO1269 prot
6	64	74.4	196	AAW25583	Htag7 protein enco
7	64	74.4	196	AAW99664	Chondrosarcoma pep
8	64	74.4	196	AAW99400	Human PRO1269 (UNQ
9	64	74.4	196	AAW66149	Protein of the inv
10	54	62.8	81	ABP08874	Human ORFX protein

11	49	57.0	190	22	ABB64581	Drosophila melanog
12	47	54.7	977	14	AAW51282	Helminth aminopect
13	46	53.5	334	22	ABG28884	Novel human diagno
14	46	53.5	432	22	AAW73613	Bacillus sp. KSM-K
15	45	52.3	75	22	AAW76272	Human colon cancer
16	45	52.3	241	23	ABB53271	Human polypeptide
17	45	52.3	368	21	AAW95963	Human healing tiss
18	45	52.3	369	22	AAE00693	Human full length
19	45	52.3	369	23	ABB53272	Human polypeptide
20	45	52.3	375	22	AAE00692	Human full length
21	45	52.3	1478	22	ABB61656	Drosophila melanog
22	44	51.2	363	22	ABG27581	Novel human diagno
23	42	48.8	110	21	AAW12437	Zea mays protein f
24	42	48.8	126	21	AAW12436	Zea mays protein f
25	42	48.8	128	21	AAW12435	Zea mays protein f
26	42	48.8	530	22	AAW72664	Murine peptidoglyc
27	42	48.8	616	16	AAW73005	Amino peptidase O12
28	42	48.8	616	16	AAW73007	Amino peptidase O12
29	42	48.8	964	12	AAW13618	C3 vegetable PEPC.
30	41	47.7	29	22	AAW83801	Human immune/haema
31	41	47.7	83	22	ABB17849	Human nervous syst
32	41	47.7	83	22	AAU20570	Human secreted pro
33	41	47.7	185	22	ABB59234	Drosophila melanog
34	41	47.7	185	22	ABB69758	Drosophila melanog
35	41	47.7	199	22	ABB11583	Human diacylglycer
36	41	47.7	447	23	ABB79169	Carassius auratus
37	41	47.7	655	22	ABG24248	Novel human diagno
38	41	47.7	736	22	AAU35700	Helicobacter pylor
39	41	47.7	736	22	AAU35880	Helicobacter pylor
40	41	47.7	830	21	AAW30503	Arabidopsis thalia
41	41	47.7	1065	22	AAE05111	Human diacylglycer
42	41	47.7	1076	22	ABG06003	Novel human diagno
43	40.5	47.1	445	20	AAW19821	B. burgdorferi ant
44	40.5	47.1	469	20	AAW19820	B. burgdorferi ant
45	40	46.5	57	22	ABG03096	Novel human diagno

## ALIGNMENTS

RESULT 1  
AAW23723  
ID AAW23723 standard; Protein; 181 AA.  
XX AC AAW23723;  
XX DT 18-FEB-1998 (first entry)  
XX DE Murine granulocyte peptide A precursor (antimicrobial MGP-A).  
XX KW Antimicrobial peptide; antibiotic; antibacterial; antifungal;  
KW fungicide; antiprotzoa; protozoacide; antiviral; virucide;  
KW murine granulocyte peptide A; MGP-A; preservative; sepsis;  
KW endotoxaemia; mouse.  
XX OS Mus musculus.  
XX FH Key Location/Qualifiers  
FT Peptide /label= Mat\_peptide  
FT /note= "MGP-A antimicrobial peptide (Claim 3)"  
XX PN WO9729765-A1.  
XX PD 21-AUG-1997.  
XX PF 13-FEB-1997; 97WO-US02218.  
XX PR 16-FEB-1996; 96US-0011834.  
XX PA (REGC ) UNIV CALIFORNIA.  
XX PI Selsted ME;

XX WPI; 1997-424753/39.  
 DR N-PSDB; AAT78510.  
 XX Antimicrobial (poly)peptide(s) from bovine and murine granulocytes -  
 PT useful therapeutically, as preservatives for food, in water  
 PT treatment and in agriculture  
 XX Claim 9; Fig 5; 56pp; English.  
 XX This protein comprises the precursor of a novel, claimed  
 CC antimicrobial peptide from murine neutrophils, designated murine  
 CC granulocyte peptide A or MGP-A (see AAW23725). Its amino acid  
 CC sequence was deduced from a cDNA clone (see AAT78510) obtained from  
 CC murine bone marrow. MGP-A and the bovine homologue, BGP-A (see  
 CC AAW23724), exhibit activity against Gram-positive and Gram-negative  
 CC bacteria, fungi and viruses, specifically *Staphylococcus aureus*,  
 CC *Escherichia coli*, *Candida albicans*, *Salmonella typhimurium* and *C.*  
 CC *neoformans* (claimed). They can be used in human or veterinary  
 CC medicine (particularly to treat disorders associated with  
 CC lipopolysaccharides, e.g. sepsis and endotoxaemia) or as  
 CC preservatives in food products or in water supplies (claimed).  
 CC They can also be applied to crops to reduce post-harvest spoilage  
 CC or expressed in transgenic plants to increase their disease  
 CC resistance. They have low immunogenicity. Carboxamidated analogues  
 CC of MGP-A and BGP-A may also be used.  
 XX  
 XX SQ Sequence 181 AA;  
 Query Match 100.0%; Score 86; DB 18; Length 181;  
 Best Local Similarity 100.0%; Pred. No. 5.5e-07;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGFLRSNYEVKGRDV 16  
 Db 144 RGFLRSNYEVKGRDV 159  
 |||||  
 RESULT 2  
 AAY00770  
 ID AAY00770 standard; Protein; 182 AA.  
 XX  
 AC AAY00770;  
 XX  
 DT 18-MAY-1999 (first entry)  
 XX  
 DE Mouse tag7 clone protein sequence.  
 XX  
 KW Tag7; tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma;  
 KW melanoma; leukaemia; apoptosis inducer; mouse.  
 XX  
 OS Mus sp.  
 XX  
 XX WO9902686-A1.  
 PN  
 XX  
 PD 21-JAN-1999.  
 XX  
 XX 10-JUL-1998; 98WO-EP04287.  
 PF  
 XX 11-JUL-1997; 97US-0893764.  
 PR  
 XX (BOEH ) BOEHRINGER INGELHEIM INT GMBH.  
 PA  
 XX Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;  
 PI  
 XX WPI; 1999-120887/10.  
 DR N-PSDB; AAX21819.  
 DR  
 XX New nucleic acid encoding tag7 - used to inhibit tumour growth and  
 PT induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and  
 PT leukaemia  
 XX  
 PS Claim 19; Fig 1; 138pp; English.

XX This sequence is the murine tag7 of the invention. Cells containing  
 CC the tag7 DNA sequence are used to express recombinant tag7. Tag7 is used  
 CC to produce and purify antibodies; to inhibit growth of mammalian tumours,  
 CC especially for treating carcinoma (of liver, ovary, breast, cervix, lung,  
 CC prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head  
 CC and neck, squamous cell carcinoma or teratocarcinoma), sarcoma (Kaposi's,  
 CC osteo- or fibro-sarcomas), melanoma or leukaemia; and as a molecular  
 CC weight marker. The tag7 polypeptide inhibits tumour growth and induces  
 CC apoptosis. The tag7 coding sequences are also useful as probes for gene  
 CC mapping and detection of tag7 gene expression, and as primers. Antibodies  
 CC against tag7 are used as reagents for detecting tag7; as an antagonist of  
 CC tag7; for isolating tag7 and therapeutically to inhibit or delay tumour  
 CC metastasis.  
 XX  
 XX SQ Sequence 182 AA;  
 Query Match 100.0%; Score 86; DB 20; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 5.5e-07;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGFLRSNYEVKGRDV 16  
 Db 145 RGFLRSNYEVKGRDV 160  
 |||||  
 RESULT 3  
 AAW23722  
 ID AAW23722 standard; Protein; 190 AA.  
 XX  
 AC AAW23722;  
 XX  
 DT 18-FEB-1998 (first entry)  
 XX  
 DE Bovine granulocyte peptide A precursor (antimicrobial BGP-A).  
 XX  
 KW Antimicrobial peptide; antibiotic; antibacterial; antifungal;  
 KW fungicide; antiprotozoa; protozoacide; antiviral; virucide;  
 KW bovine granulocyte peptide A; BGP-A; preservative; sepsis;  
 KW endotoxaemia; cattle.  
 XX  
 OS Bos taurus.  
 XX  
 XX Key Location/Qualifiers  
 FH Peptide 1..21  
 FT /label= Sig\_peptide  
 FT 22..177  
 FT /label= propeptide  
 FT Peptide 178..190  
 FT /label= Mat\_peptide  
 FT /note= "BGP-A antimicrobial peptide (Claim 2)"  
 XX  
 PN WO9729765-A1.  
 XX  
 XX 21-AUG-1997.  
 PD  
 XX 13-FEB-1997; 97WO-US02218.  
 PF  
 XX 16-FEB-1996; 96US-0011834.  
 PR  
 XX (REGC ) UNIV CALIFORNIA.  
 PA  
 XX Selsted ME;  
 XX  
 XX WPI; 1997-424753/39.  
 DR N-PSDB; AAT78509.  
 DR  
 XX Antimicrobial (poly)peptide(s) from bovine and murine granulocytes -  
 PT useful therapeutically, as preservatives for food, in water  
 PT treatment and in agriculture  
 XX  
 PS Claim 8; Fig 4; 56pp; English.

CC This protein comprises the precursor (see AAW23722) of a novel,  
 CC claimed antimicrobial peptide from bovine neutrophils, designated  
 CC bovine granulocyte peptide A or BGP-A (see AAW23724). Its amino acid  
 CC sequence was deduced from a cDNA clone (see AAW78509) obtained from  
 CC bovine bone marrow. BGP-A and the murine homologue, MGP-A (see  
 CC AAW23725), exhibit activity against Gram-positive and Gram-negative  
 CC bacteria, fungi and viruses, specifically *Staphylococcus aureus*,  
 CC *Escherichia coli*, *Candida albicans*, *Salmonella typhimurium* and *C.*  
 CC *neoformans* (claimed). They can be used in human or veterinary  
 CC medicine (particularly to treat disorders associated with  
 CC lipopolysaccharides, e.g. sepsis and endotoxaemia) or as  
 CC preservatives in food products or in water supplies (claimed).  
 CC They can also be applied to crops to reduce post-harvest spoilage  
 CC or expressed in transgenic plants to increase their disease  
 CC resistance. They have low immunogenicity. Carboxamidated  
 CC analogues of BGP-A and MGP-A may also be used.

XX  
 XX  
 SQ Sequence 190 AA;

Query Match 83.7%; Score 72; DB 18; Length 190;  
 Best Local Similarity 81.2%; Pred. No. 0.00016;  
 Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 RGFLRSNYEVKGRDV 16  
 DB 152 RGYLPNRYEVKGRDV 167  
 ||:| |||||

RESULT 4  
 AAY00771  
 ID AAY00771 standard; Protein; 191 AA.  
 AC AAY00771;  
 XX  
 XX 18-MAY-1999 (first entry)  
 DT  
 DE Human tag7 clone protein sequence.  
 KW Tag7: tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma;  
 KW melanoma; leukaemia; apoptosis inducer; human.  
 XX  
 OS Homo sapiens.  
 PN WO9502686-A1.  
 XX  
 XX 21-JAN-1999.  
 PD  
 PF 10-JUL-1998; 98WO-EP04287.  
 XX  
 XX 11-JUL-1997; 97US-0893764.  
 PR  
 PA (BOE) BOEHRINGER INGELHEIM INT GMBH.  
 XX  
 PI Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;  
 DR WPI; 1999-120887/10.  
 DR N-PSDB; AAX21820.  
 XX  
 XX New nucleic acid encoding tag7 - used to inhibit tumour growth and  
 PT induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and  
 PT leukaemia  
 XX  
 XX Claim 23; Page 126-127; 138pp; English.

CC This sequence is the human tag7 of the invention. Cells containing  
 CC the tag7 DNA sequence are used to express recombinant tag7. Tag7 is used  
 CC to produce and purify antibodies; to inhibit growth of mammalian tumours,  
 CC especially for treating carcinoma (of liver, ovary, breast, cervix, lung,  
 CC prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head  
 CC and neck, squamous cell carcinoma or teratocarcinoma), sarcoma (Kaposi's,  
 CC osteo- or fibro-sarcomas), melanoma or leukaemia; and as a molecular  
 CC weight marker. The tag7 polypeptide inhibits tumour growth and induces  
 CC apoptosis. The tag7 coding sequences are also useful as probes for gene

CC mapping and detection of tag7 gene expression, and as primers. Antibodies  
 CC against tag7 are used as reagents for detecting tag7; as an antagonist of  
 CC tag7; for isolating tag7 and therapeutically to inhibit or delay tumour  
 CC metastasis.

XX  
 XX  
 SQ Sequence 191 AA;

Query Match 74.4%; Score 64; DB 20; Length 191;  
 Best Local Similarity 75.0%; Pred. No. 0.0038;  
 Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 RGFLRSNYEVKGRDV 16  
 DB 153 QGALRSNYVLKGRDV 168  
 :||| |||||

RESULT 5  
 AAB24022  
 ID AAB24022 standard; Protein; 196 AA.  
 XX  
 AC AAB24022;  
 XX  
 XX 25-JAN-2001 (first entry)  
 DT  
 DE Human PRO1269 protein sequence SEQ ID NO:7.  
 KW Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;  
 KW identification; tumorigenesis; anticancer; detection.  
 XX  
 OS Homo sapiens.  
 PN WO200053750-A1.  
 XX  
 XX 14-SEP-2000.  
 PD  
 PF 02-DEC-1999; 99WO-US28551.  
 XX  
 XX 08-MAR-1999; 99WO-US05028.  
 PR 01-SEP-1999; 99WO-US20111.  
 PR 29-OCT-1999; 99US-0162506.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 01-DEC-1999; 99WO-US28634.  
 XX  
 XX (GETH) GENENTECH INC.  
 XX  
 XX Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WT;  
 PI  
 XX WPI; 2000-594320/56.  
 DR  
 DR N-PSDB; AAC58104.  
 XX  
 XX Antibodies specific for PRO polypeptides, used to diagnose and inhibit  
 PT the growth of tumors in mammals, and to identify inhibitors of PRO  
 PT polypeptide activity or expression -  
 XX  
 XX Claim 61; Fig 4; 226pp; English.

CC The present invention describes an antibody that binds to a human  
 CC protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;  
 CC PRO3434; PRO1927; PRO3567; PRO1295; PRO1303; PRO4344; PRO4354;  
 CC PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has  
 CC anticancer activity and can be used to diagnose tumours in mammals, by  
 CC detecting complex formation when the antibody is contacted with test  
 CC cells. Increased expression of genes encoding (I) can also be detected  
 CC to diagnose tumours. Agents which inhibit the activity of (I),  
 CC especially the antibodies, or an antisense oligonucleotide which  
 CC hybridises to genes encoding (I), can be used to inhibit tumour growth,  
 CC preferably by inducing cell death. Methods from the present invention  
 CC can be used to identify compounds which inhibit the biological activity  
 CC of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation  
 CC probes used in examples from the present invention for human PRO  
 CC sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human  
 CC PRO polynucleotide and protein sequences given in the exemplification of  
 CC the present invention.

xx  
SQ Sequence 196 AA;  
Query Match 74.4%; Score 64; DB 21; Length 196;  
Best Local Similarity 75.0%; Pred. No. 0.0039;  
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 RGLRSLNVEYKGRDV 16  
Db 158 QGALRSNYVLKGRDV 173  
RESULT 6  
AAB25583  
ID AAB25583 standard; Protein: 196 AA.  
AC AAB25583;  
DT 21-NOV-2000 (first entry)  
XX Htag7 protein encoded by human secreted protein gene #8.  
DE Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;  
KW antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;  
KW anticancer; vulnary; antiviral; antibacterial; antifungal;  
KW immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;  
KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;  
KW Crohn's disease; nephritis; hyperproliferative disorder;  
KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;  
KW melanoma; lymphoma; wound healing; human.  
XX  
OS Homo sapiens.  
XX  
PN WO200029435-A1.  
XX  
PD 25-MAY-2000.  
XX  
PF 27-OCT-1999; 99WO-US25031.  
XX  
PR 28-OCT-1998; 98US-0105971.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;  
PI Greene JM;  
XX  
DR WPI; 2000-387742/33.  
XX  
DR N-PSDB; AAA80613.  
XX  
PT Isolated nucleic acid molecules encoding human secreted proteins are  
PT used for the prevention, amelioration and treatment of autoimmune,  
PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,  
PT wounds, and infectious diseases  
XX  
PS Claim 1; Figure 34; 803pp; English.  
XX  
CC The present invention relates to 12 secreted human proteins and the  
CC nucleotide sequences encoding them. The polynucleotide sequences given  
CC in AAA80606-A80623 encode the 12 secreted protein sequences given in  
CC AAB25576-B25593. The human secreted proteins have various activities  
CC dependent on the tissues in which they are expressed. Examples of the  
CC activities of the proteins include: immunosuppressant;  
CC anti-inflammatory; antiarthritic; antirheumatic, dermatological;  
CC antiproliferative; antiarteriosclerotic; anticancer; vulnary;  
CC antiviral; antibacterial; and antifungal activity. The proteins,  
CC polypeptides, agonists and antagonists may be used to treat prevent  
CC and/or diagnose various disease, disorders and conditions examples of  
CC which include: immune disorders e.g. Addison's disease, rheumatoid  
CC arthritis, dermatitis, and multiple sclerosis; inflammatory disorders  
CC e.g. inflammatory bowel disease, Crohn's disease and nephritis;  
CC hyperproliferative disorders such as paraproteinemia and purpura;  
CC cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;  
CC cancer e.g. melanoma and lymphoma. The proteins and polynucleotide

CC sequences may also be used in wound healing and the treatment of  
CC infectious diseases. The human secreted protein gene #8 and protein  
CC sequences are represented in sequences AAA80613 and AAB25583. Sequences  
CC AAA80606-A80663 represent genes related to the secreted protein gene#8.  
XX  
SQ Sequence 196 AA;  
Query Match 74.4%; Score 64; DB 21; Length 196;  
Best Local Similarity 75.0%; Pred. No. 0.0039;  
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 RGLRSLNVEYKGRDV 16  
Db 158 QGALRSNYVLKGRDV 173  
RESULT 7  
AAY96964  
ID AAY96964 standard; Protein: 196 AA.  
XX  
AC AAY96964;  
XX  
DT 31-OCT-2000 (first entry)  
XX Chondrosarcoma peptidoglycan recognition protein-like protein.  
DE Peptidoglycan recognition protein-like protein; PGRP; PGRP-C; regulator;  
XX Chondrosarcoma; immunosuppressive; antibacterial; anti-apoptotic;  
KW tumour necrosis factor; TNF; anti-infectious; mononuclear phagocyte;  
KW inhibitor; protein co-ordinate data.  
KW  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT Protein /label= Signal\_peptide  
FT Domain 22..196  
FT /label= Mature\_protein  
FT Region 34..117  
FT /label= PGRP-like\_domain  
FT Region 1..6  
FT /label= Antigenic\_region  
FT Region 20..29  
FT /label= Antigenic\_region  
FT Region 33..43  
FT /label= Antigenic\_region  
FT Region 63..79  
FT /label= Antigenic\_region  
FT Region 99..112  
FT /label= Antigenic\_region  
FT Region 133..146  
FT /label= Antigenic\_region  
FT Region 160..165  
FT /label= Antigenic\_region  
FT Region 168..181  
FT /label= Antigenic\_region  
FT Region 190..196  
FT /label= Antigenic\_region  
PN WO200039327-A1.  
XX  
PD 06-JUL-2000.  
XX  
PF 22-DEC-1999; 99WO-US30736.  
XX  
PR 23-DEC-1998; 98US-0113809.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA Rosen CA, Ruben SM, Young PE, Olsen HS;  
PI WPI; 2000-452414/39.  
XX  
DR N-PSDB; AAA51719.

XX Polynucleotide encoding peptidoglycan recognition protein-like protein,  
 PT antibodies specific to it useful for preventing, treating conditions  
 PT e.g. endotoxemic shock and auto-immune disorders and infections in mammal  
 XX Claim 11; Fig 3; 191pp; English.

CC Novel human peptidoglycan recognition protein-like proteins (PGRP)  
 CC expressed by keratinocytes (PGRP-K), wound healing tissues (PGRP-W)  
 CC or chondrosarcoma tissue (PGRP-C) are disclosed. The PGRP DNA and  
 CC proteins are useful for preventing, treating or ameliorating a medical  
 CC condition in a mammal (claimed). PGRP is useful in augmenting the immune  
 CC system in such areas as immune recognition, antigen presentation and  
 CC immune system activation. Antibodies or antagonists directed against  
 CC these proteins may be useful in reducing or eliminating disorders  
 CC associated with tumour necrosis factor (TNF) and TNF-like cytokines,  
 CC such as endotoxemic shock and autoimmune disorders and for treating  
 CC infectious diseases including silicosis, sarcoidosis and idiopathic  
 CC pulmonary fibrosis.

XX Sequence 196 AA;

Query Match 74.4%; Score 64; DB 21; Length 196;  
 Best Local Similarity 75.0%; Pred. No. 0.0039;  
 Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 RGFLRSNYEVKGRDV 16

DB 158 QGALRSNYVLKGRDV 173

# RESULT 8

AA999400  
 ID AAY99400 standard; Protein; 196 AA.

AC AAY99400;

DT 08-AUG-2000 (first entry)

DE Human PRO1269 (UNQ639) amino acid sequence SEQ ID NO:216.

KW Human: PRO polypeptide; membrane bound protein; receptor; diagnosis;  
 KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.

XX Homo sapiens.

OS WO200012708-A2.

PN 09-MAR-2000.

PD 01-SEP-1999;

PF 99WO-US20111.

XX 01-SEP-1998; 98US-0098716.

PR 01-SEP-1998; 98US-0098749.

PR 01-SEP-1998; 98US-0098750.

PR 02-SEP-1998; 98US-0098803.

PR 02-SEP-1998; 98US-0098821.

PR 02-SEP-1998; 98US-0098843.

PR 09-SEP-1998; 98US-0099536.

PR 09-SEP-1998; 98US-0099596.

PR 09-SEP-1998; 98US-0099598.

PR 09-SEP-1998; 98US-0099602.

PR 10-SEP-1998; 98US-0099741.

PR 10-SEP-1998; 98US-0099754.

PR 10-SEP-1998; 98US-0099763.

PR 10-SEP-1998; 98US-0099792.

PR 15-SEP-1998; 98US-0100390.  
 PR 16-SEP-1998; 98US-0100584.  
 PR 16-SEP-1998; 98US-0100627.  
 PR 16-SEP-1998; 98US-0100661.  
 PR 16-SEP-1998; 98US-0100662.  
 PR 16-SEP-1998; 98US-0100664.  
 PR 17-SEP-1998; 98US-0100683.  
 PR 17-SEP-1998; 98US-0100684.  
 PR 17-SEP-1998; 98US-0100710.  
 PR 17-SEP-1998; 98US-0100711.  
 PR 17-SEP-1998; 98US-0100919.  
 PR 17-SEP-1998; 98US-0100930.  
 PR 18-SEP-1998; 98US-0100848.  
 PR 18-SEP-1998; 98US-0100849.  
 PR 18-SEP-1998; 98US-0101014.  
 PR 18-SEP-1998; 98US-0101068.  
 PR 18-SEP-1998; 98US-0101071.  
 PR 22-SEP-1998; 98US-0101279.  
 PR 23-SEP-1998; 98US-0101471.  
 PR 23-SEP-1998; 98US-0101472.  
 PR 23-SEP-1998; 98US-0101474.  
 PR 23-SEP-1998; 98US-0101475.  
 PR 23-SEP-1998; 98US-0101476.  
 PR 23-SEP-1998; 98US-0101477.  
 PR 24-SEP-1998; 98US-0101479.  
 PR 24-SEP-1998; 98US-0101738.  
 PR 24-SEP-1998; 98US-0101741.  
 PR 24-SEP-1998; 98US-0101743.  
 PR 24-SEP-1998; 98US-0101915.  
 PR 29-SEP-1998; 98US-0101916.  
 PR 29-SEP-1998; 98US-0102207.  
 PR 29-SEP-1998; 98US-0102240.  
 PR 29-SEP-1998; 98US-0102307.  
 PR 29-SEP-1998; 98US-0102330.  
 PR 30-SEP-1998; 98US-0102331.  
 PR 30-SEP-1998; 98US-0102484.  
 PR 30-SEP-1998; 98US-0102487.  
 PR 30-SEP-1998; 98US-0102570.  
 PR 30-SEP-1998; 98US-0102571.  
 PR 01-OCT-1998; 98US-0102684.  
 PR 01-OCT-1998; 98US-0102687.  
 PR 02-OCT-1998; 98US-0102965.  
 PR 06-OCT-1998; 98US-0103258.  
 PR 06-OCT-1998; 98US-0103259.  
 PR 07-OCT-1998; 98US-0103314.  
 PR 07-OCT-1998; 98US-0103315.  
 PR 07-OCT-1998; 98US-0103328.  
 PR 07-OCT-1998; 98US-0103395.  
 PR 07-OCT-1998; 98US-0103396.  
 PR 08-OCT-1998; 98US-0103401.  
 PR 08-OCT-1998; 98US-0103633.  
 PR 08-OCT-1998; 98US-0103678.  
 PR 08-OCT-1998; 98US-0103679.  
 PR 14-OCT-1998; 98US-0103711.  
 PR 20-OCT-1998; 98US-0104257.  
 PR 20-OCT-1998; 98US-0104987.  
 PR 20-OCT-1998; 98US-0105000.  
 PR 21-OCT-1998; 98US-0105002.  
 PR 22-OCT-1998; 98US-0105104.  
 PR 22-OCT-1998; 98US-0105169.  
 PR 26-OCT-1998; 98US-0105266.  
 PR 26-OCT-1998; 98US-0105693.  
 PR 27-OCT-1998; 98US-0105807.  
 PR 27-OCT-1998; 98US-0105881.  
 PR 27-OCT-1998; 98US-0105882.  
 PR 28-OCT-1998; 98US-0106062.  
 PR 28-OCT-1998; 98US-0106023.  
 PR 28-OCT-1998; 98US-0106029.  
 PR 28-OCT-1998; 98US-0106030.  
 PR 28-OCT-1998; 98US-0106032.  
 PR 28-OCT-1998; 98US-0106033.  
 PR 28-OCT-1998; 98US-0106178.



XX Homo sapiens.  
OS WO200192523-A2.  
PN 06-DEC-2001.  
PD 29-MAY-2001; 2001WO-US10836.  
XX 30-MAY-2000; 2000US-206132P.  
XX 29-AUG-2000; 2000US-228716P.  
XX (CURA-) CURAGEN CORP.  
PA Shimkets RA, Leach MD;  
PI WPI; 2002-106308/14.  
PI N-PSDB; ABN24626.  
DR Novel human polypeptides and polynucleotides useful for diagnosing,  
DR preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders  
PT  
XX Disclosure; SEQ ID 17730; 1037pp; English.  
XX  
XX The present invention describes substantially purified human proteins  
XX (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
XX in the specification)). ABN15762 to ABN27252 encode the human ORFX  
XX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
XX treating or preventing a pathology associated with an ORFX-associated  
XX disorder in humans, and in the manufacture of a medicament for treating a  
XX syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
XX sequences can be used in gene therapy. ORFX sequences can be used in the  
XX treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
XX psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
XX osteoarthritis, neurodegenerative disorders, disorders related to organ  
XX transplantation, cardiovascular diseases, diabetes mellitus, systemic  
XX lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
XX storage disease, various immune deficiencies and disorders, infectious  
XX diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
XX arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
XX disease and autoimmune inflammatory eye disease. ORFX proteins are also  
XX useful for treating burns, incisions, ulcers, for treating osteoporosis,  
XX bone degenerative disorders, or periodontal disease, and for gut  
XX protection or regeneration and treatment of lung or liver fibrosis,  
XX reperfusion injury in various tissues and conditions resulting from  
XX systemic cytokine damage.  
XX N.B. The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 81 AA;  
XX  
Query Match 62.8%; Score 54; DB 23; Length 81;  
Best Local Similarity 64.3%; Pred. No. 0.078;  
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
QY 1 RGLFRSNVEYKGRH 14  
DB 61 RGFIRGNYEIKENR 74  
|||:| |||:| :|  
RESULT 11  
ABB64581  
ID ABB64581 standard; Protein; 190 AA.  
XX ABB64581;  
AC  
XX 26-MAR-2002 (first entry)  
DT  
XX Drosophila melanogaster polypeptide SEQ ID NO 20535.  
DE  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW

KW pharmaceutical.  
XX Drosophila melanogaster.  
OS WO200171042-A2.  
PN 27-SEP-2001.  
PD 23-MAR-2001; 2001WO-US09231.  
XX 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
XX (PEKE ) PE CORP NY.  
PA Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
PI N-PSDB; ABL08684.  
DR New isolated nucleic acid detection reagent for detecting 1000 or more  
DR genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions  
PT  
XX Disclosure; SEQ ID NO 20535; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins  
XX (ABB57737-ABB72072).  
XX The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 190 AA;  
XX  
Query Match 57.0%; Score 49; DB 22; Length 190;  
Best Local Similarity 57.1%; Pred. No. 1.5;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 1 RGLFRSNVEYKGRH 14  
DB 152 RGYLKDNVTLFGRH 165  
|||:| |||:| :|  
RESULT 12  
AAR51282  
ID AAR51282 standard; Protein; 977 AA.  
XX AAR51282;  
AC  
XX 31-MAY-1994 (first entry)  
DT  
XX Helminth aminopeptidase H110D variant encoded by the H11-1 gene.  
DE  
XX Helminth; aminopeptidase; H11-3; H11-2; H11-1; integral membrane;  
KW alpha-amino acyl peptide hydrolase (microsomal); multigene family;  
KW antigen; vaccine; parasite; human; virus; microbe.  
XX Haemonchus contortus.  
XX WO9323542-A.  
PN 25-NOV-1993.  
PD  
XX 07-MAY-1993; 93WO-GB00943.  
PF  
XX 08-MAY-1992; 92GB-0009993.  
PR  
XX

PA (AGRI-) AGRIC & FOOD RES COUNCIL.  
XX  
PI Graham M, Knox DP, Munn EA, Newton SE, Oliver JJ;  
PI Smith TS;  
XX  
DR WPI; 1993-386574/48.  
DR N-PSDB; AAQ52491.  
XX  
PT New DNA encoding amino-peptidase from Haemonchus contortus and  
PT derived peptide(s) - useful in protective vaccines, for new  
PT vectors, transformed cells and oligosaccharide, for incorporation  
PT in virus or microbe  
XX  
PS Claim 6; Page 78-80; 137pp; English.  
XX  
CC The sequences given in AA51280-82 represent variants of the helminth  
CC aminopeptidase H110D encoded by the genes H11-3, -2 and -1 respectively.  
CC H110D is a protein doublet which shows homology to a family of integral  
CC membrane aminopeptidases. The differences between the coding sequences  
CC and the translated products, can be attributed to different mRNAs of  
CC the multigene family, and also to different variants of the H110D-  
CC encoding sequence being present at different stages of the life  
CC cycle, or in strains differing in geographical origin. Antigenic  
CC fragments of these aminopeptidases may be used in vaccines to stimulate  
CC immune response against helminth parasites in humans or other animals.  
CC The DNA sequences encoding these proteins may be incorporated into a  
CC virus or microbe and used in a similar manner.  
XX  
SQ Sequence 977 AA;  
Query Match 54.7%; Score 47; DB 14; Length 977;  
Best Local Similarity 50.0%; Pred. No. 22;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 RGFLRSNYEVKGRDV 16  
| | | | | | | | | |  
Db 630 RAFCRSNDANGWRNI 645  
RESULT 13  
ABG26884  
ID ABG26884 standard; Protein; 334 AA.  
XX  
AC ABG26884;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #26875.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS91071.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess

PT biodiversity -  
XX Claim 20; SEQ ID No 57243; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 334 AA;  
Query Match 53.5%; Score 46; DB 22; Length 334;  
Best Local Similarity 58.3%; Pred. No. 9.7;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
Qy 4 LRSNVEVKGHRD 15  
| | | | | | | | | |  
Db 292 LRDNFDIVGHRD 303  
RESULT 14  
AAB73613  
ID AAB73613 standard; Protein; 432 AA.  
XX  
AC AAB73613;  
XX  
DT 10-AUG-2001 (first entry)  
XX  
DE Bacillus sp. KSM-KP43 origin of replication-associated protein.  
XX  
KW Origin of replication-associated protein; ori; plasmid construction;  
KW expression vector; recombinant expression.  
XX  
OS Bacillus sp. KSM-KP43.  
XX  
PN JP2001069986-A.  
XX  
PD 21-MAR-2001.  
XX  
PF 03-SEP-1999; 99JP-0250690.  
XX  
PR 03-SEP-1999; 99JP-0250690.  
XX  
PA (KAOS ) KAO CORP.  
XX  
DR WPI; 2001-297392/31.  
DR N-PSDB; AAH23948.  
XX  
PT A plasmid vector used for the preparation of a useful protein or a  
PT useful peptide -  
XX  
PS Claim 2; Page 6-7; 9pp; Japanese.  
XX  
CC The invention relates to a plasmid vector containing a 2.2 kb genomic  
CC fragment from Bacillus sp. KSM-KP43 (AAH23948) which contains the  
CC origin of replication (ori). This vector can be used for the



CC recombinant expression of a protein or peptide in a host  
CC microorganism. The invention also relates to such recombinant  
CC plasmids and host cells. In the exemplifications of the invention,  
CC a Bacillus sp. KSM-64 cellulase gene was inserted into a vector of  
CC the invention containing the Bacillus sp. KSM-KP43 ori and a  
CC tetracycline resistance gene. The present sequence represents  
CC a protein encoded by the 2.2 kb origin of replication-containing  
CC Bacillus sp. KSM-KP43 genomic fragment.  
XX  
SQ Sequence 432 AA;  
Query Match 53.5%; Score 46; DB 22; Length 432;  
Best Local Similarity 70.0%; Pred. No. 13;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
Oy 7 NYEVKGHRDV 16  
||:|||||:  
Db 247 NYQVKGHREL 256  
||:|||||:  
RESULT 15  
AAG76272  
ID AAG76272 standard; Protein; 75 AA.  
XX  
AC  
XX AAG76272;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE Human colon cancer antigen protein SEQ ID NO:7036.  
XX  
DE Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma.  
XX  
XX Homo sapiens.  
XX  
XX WO200122920-A2.  
XX  
XX 05-APR-2001.  
XX  
XX 28-SEP-2000; 2000WO-US26524.  
XX  
XX 29-SEP-1999; 99US-0157137.  
XX 03-NOV-1999; 99US-0163280.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Ruben SM, Barash SC, Birse CE, Rosen CA;  
PI  
XX  
XX WPI: 2001-235357/24.  
XX N-PSDB; AAH35677.  
XX  
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
XX useful for preventing, diagnosing and/or treating colorectal cancers -  
XX  
XX Claim 11; Page 8473-8475; 9803pp; English.  
XX  
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
XX cancer-associated nucleic acid molecules (N) and proteins (P), where  
XX the proteins are collectively known as colon cancer antigens. The colon  
XX cancer antigens have cytostatic activity and can be used in gene  
XX therapy and vaccine production. N and P may be used in the prevention,  
XX diagnosis and treatment of diseases associated with inappropriate P  
XX expression. For example, N and P may be used to treat disorders  
XX associated with decreased expression by rectifying mutations or deletions  
XX in a patient's genome that affect the activity of P by expressing  
XX inactive proteins or to supplement the patient's own production of P.  
XX Additionally, N may be used to produce the colon cancer-associated Ps,  
XX by inserting the nucleic acids into a host cell and culturing the cell  
XX to express the proteins. N and P can be used in the prevention, diagnosis  
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
XX and AAG77789 represent sequences used in the exemplification of the  
XX present invention.  
XX  
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were

CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922..  
XX  
SQ Sequence 75 AA;  
Query Match 52.3%; Score 45; DB 22; Length 75;  
Best Local Similarity 40.0%; Pred. No. 2.6;  
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
Oy 1 RGFLRSNYEVKGHRD 15  
|: : :||: ||:|  
Db 8 RNYTKHSYEISGHQD 22  
Search completed: November 4, 2002, 20:34:14  
Job time : 85 secs

GenCore version 5.1.3

Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 4, 2002, 20:33:26 ; Search time 199 Seconds  
(without alignments)  
1.115 Million cell updates/sec

Title: US-09-462-625-2\_COPY\_145\_160

Perfect score: 86

Sequence: 1 RGFLRSYEVKGRDV 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 90412 seqs, 13869272 residues

Total number of hits satisfying chosen parameters: 90412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pap.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pap.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pap.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pap.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pap.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pap.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pap.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pap.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pap.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pap.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pap.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pap.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pap.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	47.7	131	10	US-09-771-161A-148
2	41	47.7	736	10	US-09-815-242-11293
3	41	47.7	736	10	US-09-815-242-11473
4	41	47.7	757	10	US-09-725-735A-20
5	41	47.7	1065	10	US-09-771-161A-239
6	37	43.0	101	10	US-09-764-877-1470
7	37	43.0	358	10	US-09-864-761-48087
8	36	41.9	26	10	US-09-864-761-38500
9	36	41.9	407	10	US-09-815-242-5348
10	36	41.9	444	10	US-09-815-242-12376
11	36	41.9	732	10	US-09-994-485-8
12	36	41.9	777	10	US-09-925-301-1115
13	35	40.7	62	10	US-09-864-761-38482
14	35	40.7	457	10	US-09-898-570-24
15	35	40.7	558	10	US-09-815-242-5645
16	35	40.7	605	10	US-09-815-242-12658
17	35	40.7	1457	10	US-09-772-316-2
18	35	40.7	1460	10	US-09-815-242-13668
19	34	39.5	35	10	US-09-864-761-38250

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

34

34

34

34

34

34

34

34

34

34

34

34

34

34

34

34

34

34

34

34

34

34

34

34

34

34

88

137

137

137

451

451

451

451

451

451

451

451

451

451

451

451

451

451

451

451

451

451

451

451

451

451

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

## ALIGNMENTS

## RESULT 1

US-09-771-161A-148

; Sequence 148, Application US/09771161A

; Patent No. US20020110811A1

; GENERAL INFORMATION:

; APPLICANT: LEVINE, et al.

; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES

; FILE REFERENCE: 802620-2005.1

; CURRENT APPLICATION NUMBER: US/09/771,161A

; CURRENT FILING DATE: 2001-01-26

; PRIOR APPLICATION NUMBER: 09/724, 676

; PRIOR FILING DATE: 2000-11-28

; PRIOR APPLICATION NUMBER: 136776

; PRIOR FILING DATE: 2000-06-15

; PRIOR APPLICATION NUMBER: 135619

; PRIOR FILING DATE: 2000-04-12

; NUMBER OF SEQ ID NOS: 273

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 148

; LENGTH: 131

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-771-161A-148

Query Match

Best Local Similarity 47.7%; Score 41; DB 10; Length 131;

Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY

5 RSNYEYKGRDV 16

| | | | |

Db 116 RQNKVIGHEDL 127

## RESULT 2

US-09-815-242-11293

; Sequence 11293, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

Sequence 33, Appl  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 412, App  
Sequence 4, Appli  
Sequence 6, Appli  
Sequence 8, Appli  
Sequence 10, Appli  
Sequence 6, Appli  
Sequence 8, Appli  
Sequence 10, Appli  
Sequence 2, Appli  
Sequence 128, App  
Sequence 448, App  
Sequence 10025, A  
Sequence 196, App  
Sequence 33856, A  
Sequence 1406, Ap  
Sequence 971, App  
Sequence 1200, Ap  
Sequence 1334, Ap  
Sequence 41, Appl  
Sequence 58, Appl  
Sequence 5354, Ap

```
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11293
; LENGTH: 736
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11293

Query Match 47.7%; Score 41; DB 10; Length 736;
Best Local Similarity 55.0%; Pred. No. 20;
Matches 11; Conservative 2; Mismatches 3; Indels 4; Gaps 2;

QY 1 RGFLRSNYEV---KGH-RDV 16
: || |||| |||| ||| |||
Db 17 KNFLDKNYEVASKGHRDL 36

RESULT 3
US-09-815-242-11473
; Sequence 11473, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
```

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11473
; LENGTH: 736
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11473

Query Match 47.7%; Score 41; DB 10; Length 736;
Best Local Similarity 55.0%; Pred. No. 20;
Matches 11; Conservative 2; Mismatches 3; Indels 4; Gaps 2;

QY 1 RGFLRSNYEV---KGH-RDV 16
: || |||| |||| ||| |||
Db 17 KNFLDKNYEVASKGHRDL 36

RESULT 4
US-09-725-735A-20
; Sequence 20, Application US/09725735A
; Patent No. US20020031797A1
; GENERAL INFORMATION:
; APPLICANT: RUSING, MATTHIAS
; APPLICANT: SCHWEINS, THOMAS
; APPLICANT: DRESLER, PETRA
; APPLICANT: STOCK, WOLFGANG
; APPLICANT: KIY, THOMAS
; TITLE OF INVENTION: NOVEL NUCLEIC ACID ISOLATED FROM TETRAHYMENA WHICH
; TITLE OF INVENTION: CODES FOR A TRITERPENOID CYCLASE, ITS PRODUCTION AND
; TITLE OF INVENTION: USE
; FILE REFERENCE: 25426.0001
; CURRENT APPLICATION NUMBER: US/09/725,735A
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: DE 199 57 889.3
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 757
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-725-735A-20

Query Match 47.7%; Score 41; DB 10; Length 757;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 RGFLRSNYEVKGHRDV 16
||||| : ||| |
Db 48 RGFLDNRFVKGCSDL 63

RESULT 5
US-09-771-161A-239
; Sequence 239, Application US/09771161A
; Patent No. US20020110811A1
; GENERAL INFORMATION:
; APPLICANT: LEVINE, et al.
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
; FILE REFERENCE: 802620-2005.1
; CURRENT APPLICATION NUMBER: US/09/771,161A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/724,676
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 136776
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 135619
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 239
; LENGTH: 1065
; TYPE: PRT
; ORGANISM: Homo sapiens
```

US-09-771-161A-239

Query Match 47.7%; Score 41; DB 10; Length 1065;  
Best Local Similarity 58.3%; Pred. No. 30;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 5 RSNEYVKGRD 16  
I I I I I I I  
Db 1050 RQNKVIGHEDL 1061

## RESULT 6

US-09-764-877-1470  
; Sequence 1470, Application US/09764877  
; Patent No. US20020147140A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC005  
; CURRENT APPLICATION NUMBER: US/09/764,877  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 4031  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1470  
; LENGTH: 101  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-877-1470

Query Match 43.0%; Score 37; DB 10; Length 101;  
Best Local Similarity 54.5%; Pred. No. 11;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 6 SNEYVKGRD 16  
I I I I I I I  
Db 80 SSYELQGHLEV 90

## RESULT 7

US-09-864-761-48087  
; Sequence 48087, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 48087  
; LENGTH: 358  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AF117829.1  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9  
; OTHER INFORMATION: SWISSPROT HIT: Q9Y236, EVALUE 0.00e+00  
; OTHER INFORMATION: EST\_HUMAN HIT: AW995310.1, EVALUE 3.00e-94  
US-09-864-761-48087

Query Match 43.0%; Score 37; DB 10; Length 358;  
Best Local Similarity 28.6%; Pred. No. 44;  
Matches 4; Conservative 8; Mismatches 2; Indels 0; Gaps 0;

Qy 3 FLRSNEYVKGRD 16  
I I I I I I I  
Db 83 FIKRWWEIRGYORI 96

## RESULT 8

US-09-864-761-38500  
; Sequence 38500, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 38500
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005288.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 28
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 23
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 25
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 25
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 23
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 23
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 31
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 24
; OTHER INFORMATION: EST_HUMAN HIT: BF433100.1, EVALUATE 2.00e-03
US-09-864-761-38500

Query Match 41.9%; Score 36; DB 10; Length 26;
Best Local Similarity 66.7%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RGFLRSNYEVKG 12
   |||||
Db 15 RGLRYNVVCG 26

RESULT 9
US-09-815-242-5348
; Sequence 5348, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12376
; LENGTH: 444
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12376

Query Match 41.9%; Score 36; DB 10; Length 444;
Best Local Similarity 56.2%; Pred. No. 82;
Matches 9; Conservative 3; Mismatches 2; Indels 2; Gaps 2;

Qy 2 GFLRS-NYEVKGHRDV 16
   |||||
Db 81 GFLRTVNYIS-KGEKDI 95

RESULT 11
US-09-994-485-8
; Sequence 8, Application US/09994485
```

```
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5348
; LENGTH: 407
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5348

Query Match 41.9%; Score 36; DB 10; Length 407;
Best Local Similarity 56.2%; Pred. No. 75;
Matches 9; Conservative 3; Mismatches 2; Indels 2; Gaps 2;

Qy 2 GFLRS-NYEVKGHRDV 16
   |||||
Db 58 GFLRTVNYIS-KGEKDI 72

RESULT 10
US-09-815-242-12376
; Sequence 12376, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12376
; LENGTH: 444
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12376

Query Match 41.9%; Score 36; DB 10; Length 444;
Best Local Similarity 56.2%; Pred. No. 82;
Matches 9; Conservative 3; Mismatches 2; Indels 2; Gaps 2;

Qy 2 GFLRS-NYEVKGHRDV 16
   |||||
Db 81 GFLRTVNYIS-KGEKDI 95

RESULT 11
US-09-994-485-8
; Sequence 8, Application US/09994485
```

Patent No. US20020142429A1  
GENERAL INFORMATION:  
APPLICANT: Ryzanov, Alexey G.  
Hait, William N.  
Pavur, Karen S.  
TITLE OF INVENTION: ELONGATION FACTOR-2 KINASE (EF-2 KINASE)  
AND METHODS OF USE THEREFOR  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David A. Jackson, Esq.  
STREET: 411 Hackensack Ave, Continental Plaza, 4th  
Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/994,485  
FILING DATE: 27-Nov-2002 US20020142429A1-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/914,999  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 601-1-078  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 732 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Dictyostelium discoideum  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-994-485-8  
Query Match 41.9%; Score 36; DB 10; Length 732;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 5 RSNEYVKGH 13  
I I I I I I  
Db 613 RCNVTLKGH 621  
RESULT 12  
US-09-925-301-1115  
Sequence 1115, Application US/09925301  
Patent No. US20020052308A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA106  
CURRENT APPLICATION NUMBER: US/09/925,301  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05882  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1694  
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1115  
LENGTH: 777  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (5)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (9)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (14)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (21)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (32)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-301-1115  
Query Match 41.9%; Score 36; DB 10; Length 777;  
Best Local Similarity 42.9%; Pred. No. 1.5e+02;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
QY 3 FLRSNEYVKGHRDV 16  
I I I I I I I I I I  
Db 413 FIRSEYKRGFOEV 426  
RESULT 13  
US-09-864-761-38482  
Sequence 38482, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 38482  
LENGTH: 62  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC004049.1  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1  
OTHER INFORMATION: EST\_HUMAN HIT: AW95139.1, EVALUE 4.00e-30  
OTHER INFORMATION: SWISSPROT HIT: P90648, EVALUE 3.00e-05  
US-09-864-761-38482

Query Match 40.7%; Score 35; DB 10; Length 62;  
Best Local Similarity 60.0%; Pred. No. 14;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 7 NYEVKGHRDV 16  
||| ||||  
DB 48 SYEFSNRDV 57

RESULT 14  
US-09-898-570-24  
Sequence 24, Application US/09898570  
Patent No. US20020123612A1  
GENERAL INFORMATION:  
APPLICANT: GERLACH, VALERIE L.  
APPLICANT: ELLERMAN, KAREN  
APPLICANT: MACDOUGALL, JOHN R.  
APPLICANT: SMITHSON, GLENDA  
TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND  
TITLE OF INVENTION: METHODS OF USING THE SAME  
FILE REFERENCE: 15966-776CIP  
CURRENT APPLICATION NUMBER: US/09/898,570  
PRIOR FILING DATE: 2001-07-03  
PRIOR APPLICATION NUMBER: 60/198,293  
PRIOR FILING DATE: 2000-04-19  
PRIOR APPLICATION NUMBER: 60/198,645  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: 60/210,809  
PRIOR FILING DATE: 2000-06-09  
PRIOR APPLICATION NUMBER: 60/199,476  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/200,025  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/224,610  
PRIOR FILING DATE: 2000-08-11  
PRIOR APPLICATION NUMBER: 60/200,024  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/199,880  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/218,591  
PRIOR FILING DATE: 2000-07-17  
PRIOR APPLICATION NUMBER: 60/271,814  
PRIOR FILING DATE: 2001-02-27  
PRIOR APPLICATION NUMBER: 60/215,855  
PRIOR FILING DATE: 2000-07-03  
PRIOR APPLICATION NUMBER: 09/839,446

PRIOR FILING DATE: 2001-04-19  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 24  
LENGTH: 457  
TYPE: PRT  
ORGANISM: Unknown Organism  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: POLYX  
OTHER INFORMATION: AC016030\_A.0.82  
US-09-898-570-24

Query Match 40.7%; Score 35; DB 10; Length 457;  
Best Local Similarity 46.2%; Pred. No. 1.2e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 GFLRSNYEVKGHR 14  
| | | | | | | |  
DB 128 GVSTRNHEVEGKH 140

RESULT 15  
US-09-815-242-5645  
Sequence 5645, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haseibeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5645  
LENGTH: 558  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5645

Query Match 40.7%; Score 35; DB 10; Length 558;  
Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GFLRSNYEV 10  
||| ||| ||  
DB 344 GFVRSNNEV 352

Search completed: November 4, 2002, 20:41:26  
Job time : 199 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 4, 2002, 20:31:22 ; Search time 50 Seconds  
(without alignments)  
30.763 Million cell updates/sec

Title: US-09-462-625-2\_COPY\_145\_160  
Perfect score: 86  
Sequence: 1 RGFLRSYEVKGHRDV 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	52.3	471	2 S61333	Iga-specific metal
2	44.5	51.7	504	1 G71248	tldd homolog PH024
3	43	50.0	960	2 S18240	phosphoenolpyruvat
4	43	50.0	965	2 T09846	phosphoenolpyruvat
5	43	50.0	967	2 JH0667	phosphoenolpyruvat
6	42	48.8	787	2 A72230	endopeptidase La (
7	42	48.8	960	2 JH0381	phosphoenolpyruvat
8	42	48.8	960	2 S31159	phosphoenolpyruvat
9	42	48.8	964	1 Q1NT	phosphoenolpyruvat
10	42	48.8	966	2 S18318	phosphoenolpyruvat
11	42	48.8	966	2 S2853	phosphoenolpyruvat
12	42	48.8	967	2 S25081	phosphoenolpyruvat
13	41	47.7	736	2 B71972	phosphoenolpyruvat
14	41	47.7	736	2 D64534	DNA topoisomerase
15	41	47.7	830	2 T01058	DNA topoisomerase
16	40.5	47.1	318	2 G75279	hypothetical prote
17	40.5	47.1	469	2 B70201	hypothetical prote
18	40.5	47.1	652	2 T19783	hypothetical prote
19	40	46.5	328	2 T06547	hypothetical prote
20	40	46.5	799	2 F83549	probable ATP-depen
21	40	46.5	806	1 E70131	endopeptidase La (
22	40	46.5	809	2 S32899	feric-pseudobacti
23	40	46.5	819	1 C71527	endopeptidase La (
24	40	46.5	819	1 B72128	endopeptidase La (
25	40	46.5	819	2 E86494	Lon ATP-dependent
26	40	46.5	819	2 E81681	proteinasin, Lon fa
27	40	46.5	966	1 Q1X1	phosphoenolpyruvat
28	40	46.5	966	2 S28614	phosphoenolpyruvat
29	40	46.5	1025	2 I59331	thyrotropin-releas

30 40 46.5 1406 2 T17429 gag-pol polyprotei  
31 39 45.3 126 2 AB2330 30S ribosomal prot  
32 39 45.3 490 2 T36920 hypothetical prote  
33 39 45.3 651 2 F64417 probable ATP-depen  
34 39 45.3 782 2 T32155 hypothetical prote  
35 39 45.3 813 1 D70176 endopeptidase La ( 36 39 45.3 966 2 S37072 phosphoenolpyruvat 37 39 45.3 967 2 S25082 phosphoenolpyruvat 38 39 45.3 1291 2 E82325 conserved hypotet 39 38.5 44.8 341 2 AD0291 purine nucleotide 40 38.5 44.8 1566 2 T20058 hypothetical prote 41 38 44.2 264 2 T10530 thyroxine deiodina 42 38 44.2 288 2 T14863 porin Mipl - Norwa 43 38 44.2 338 2 T26399 hypothetical prote 44 38 44.2 370 1 Q0HS4C GTP-binding protei 45 38 44.2 370 2 C84267 GTP-binding protei |

## ALIGNMENTS

## RESULT 1

S61333  
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Haemophilus influenzae (st  
N; Alternate names: Iga1 protease  
C:Species: Haemophilus influenzae  
A:Variety: HK284  
C:Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 08-Dec-2000  
C:Accession: S61333  
R:Lonholt, H.; Poulsen, K.; Mogens, K.  
Mol. Microbiol. 15, 495-506, 1995  
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss  
A:Reference number: S61314; MUID:95302961; PMID:7783620  
A:Accession: S61333  
A:Status: Preliminary; nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-471 <LOM>  
A:Cross-references: EMBL:X82487; NID:g773205; PIDN:CAA57870.1; PID:g773206  
A:Experimental source: strain HK284  
C:Genetics:  
A:Gene: iga  
C:Superfamily: Iga-specific metalloendopeptidase  
C:Keywords: hydrolase; metalloproteinase

Query Match 52.3%; Score 45; DB 2; Length 471;  
Best Local Similarity 61.5%; Pred. No. 9.8;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 FLRSNYEVKGHRD 15  
| : ||||| |  
DB 357 FFKNGYEVKGKTD 369

## RESULT 2

G71248  
tldd homolog PH0246 - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C:Date: 29-Jan-1999 #sequence\_revision 29-Jan-1999 #text\_change 21-Jul-2000  
C:Accession: G71248  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hino, Y.; Yamamoto, S.; Se  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu  
DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil  
A:Reference number: A71000; MUID:98344137; PMID:9679194  
A:Accession: G71248  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-504 <KAW>  
A:Cross-references: GB:AP000001; NID:g3236128; PIDN:BAA29318.1; PID:g3256635  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interlm accession for a sequence replaced by GenBa  
C:Genetics:  
A:Gene: PH0246



C:Superfamily: Escherichia coli tldD protein

Query Match 51.7%; Score 44.5; DB 1; Length 504;  
Best Local Similarity 62.5%; Pred. No. 13;  
Matches 10; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 2 GFLRSNVEKGH-RDV 16

I:I I:I:I:I I:I

Db 436 GYLVEGEIKGLRDV 451

RESULT 3

S18240

phosphoenolpyruvate carboxylase (EC 4.1.1.31) - sorghum

C:Species: Sorghum bicolor (sorghum)

C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 22-Jun-1999

C:Accession: S18240

R:Lepiniec, L.; Santi, S.; Keryer, E.; Amiet, V.; Vidal, J.; Gadai, P.; Cretin, C.

Plant Mol. Biol. 17, 1077-1079, 1991

A:Title: Complete nucleotide sequence of one member of the Sorghum phosphoenolpyruvate

A:Reference number: S18240; MUID:92032766; PMID:1840686

A:Accession: S18240

A:Molecule type: DNA

A:Residues: 1-960 <LEP>

A:Cross-references: EMBL:X59925; NID:g22592; PIDN:CAA42549.1; PID:g22593

A:Note: the authors translated the codon TTG for residue 395 as Phe, AAT for residue 698

C:Genetics:

A:Introns: 51/3; 182/3; 210/3; 284/3; 319/3; 348/3; 400/3; 733/3; 862/3

C:Superfamily: phosphoenolpyruvate carboxylase

C:Keywords: carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase

Query Match

Best Local Similarity 50.0%; Score 43; DB 2; Length 960;

Matches 10; Conservative 2; Mismatches 1; Indels 8; Gaps 1;

QY 4 LRSNVE-----VKGHRDV 16

I:I:I I:I:I I:I:I:I

Db 849 LRANYEETKLLQVAGHRDL 869

RESULT 4

T09846

phosphoenolpyruvate carboxylase (EC 4.1.1.31) 1 - upland cotton

C:Species: Gossypium hirsutum (upland cotton)

C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 24-Nov-1999

C:Accession: T09846

R:Vojdani, F.; Wilkins, T.A.

submitted to the EMBL Data Library, June 1997

A:Reference number: Z16883

A:Accession: T09846

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-965 <VO>

A:Cross-references: EMBL:AF008939; NID:g2266946; PID:g2266947

A:Experimental source: strain Acala SJ-2; fiber

C:Genetics:

A:Gene: PEPC1

C:Function: <PEPC>

A:Description: catalyzes incorporation of a carbon dioxide molecule into phosphoenolpyru

A:Pathway: tricarboxylic acid cycle

C:Function: <CDP>

A:Description: catalyzes the carboxylation (by carbon dioxide) of phosphoenolpyruvate to

A:Pathway: carbon dioxide fixation

C:Superfamily: phosphoenolpyruvate carboxylase

C:Keywords: carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase; phosphoprotein

Query Match

Best Local Similarity 50.0%; Score 43; DB 2; Length 965;

Matches 10; Conservative 2; Mismatches 1; Indels 8; Gaps 1;

QY 4 LRSNVE-----VKGHRDV 16

I:I:I I:I:I I:I:I:I

Db 854 LRANYEETKRLVQVAGHRDL 874

RESULT 5

JH0667

phosphoenolpyruvate carboxylase (EC 4.1.1.31) C3-form - maize

C:Species: Zea mays (maize)

C:Date: 09-Oct-1992 #sequence\_revision 09-Oct-1992 #text\_change 22-Jun-1999

C:Accession: JH0667

R:Kawamura, T.; Shigesada, K.; Toh, H.; Okumura, S.; Yanagisawa, S.; Izui, K.

J. Biochem. 112, 147-154, 1992

A:Title: Molecular evolution of phosphoenolpyruvate carboxylase for C4 photosynthesis

tic function.

A:Reference number: JH0667; MUID:93054411; PMID:1429504

A:Accession: JH0667

A:Molecule type: mRNA

A:Residues: 1-967 <KAW>

A:Cross-references: GB:X61489; NID:g429148; PIDN:CAA43709.1; PID:g429149

A:Experimental source: root

C:Comment: This enzyme catalyzes the carboxylation of phosphoenolpyruvate to form oxa

C:Superfamily: phosphoenolpyruvate carboxylase

C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match

Best Local Similarity 50.0%; Score 43; DB 2; Length 967;

Matches 10; Conservative 2; Mismatches 1; Indels 8; Gaps 1;

QY 4 LRSNVE-----VKGHRDV 16

I:I:I I:I:I I:I:I:I

Db 856 LRANYEETKLLQVAGHRDL 876

RESULT 6

A72230

endopeptidase La (EC 3.4.21.53) - Thermotoga maritima (strain MSB8)

N:Contains: adenosinetriphosphatase (EC 3.6.1.3)

C:Species: Thermotoga maritima

C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: A72230

R:Nelson, K.E.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,

C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: A72230

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-787 <ARN>

A:Cross-references: GB:AE001806; GB:AE000512; NID:g4982196; PIDN:AAD36700.1; PID:g498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1633

C:Superfamily: ATP-dependent serine proteinase La

C:Keywords: ATP; hydrolase; serine proteinase

Query Match

Best Local Similarity 48.8%; Score 42; DB 2; Length 787;

Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 GFLRSNVEKGHR 14

I:I I:I I:I:I:I

Db 404 GGLRDEAEIKGHR 416

RESULT 7

JH0381

phosphoenolpyruvate carboxylase (EC 4.1.1.31) - sorghum

C:Species: Sorghum bicolor

C:Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 18-Jul-2001

C:Accession: JH0381; S16455

R:Cretin, C.; Santi, S.; Keryer, E.; Lepiniec, L.; Tagu, D.; Vidal, J.; Gadai, P.

Gene 99, 87-94, 1991

A:Title: The phosphoenolpyruvate carboxylase gene family of Sorghum: promoter structu

A:Reference number: JH0381; MUID:91216449; PMID:2022326  
A:Accession: JH0381

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-960 <CRE>

A:Cross-references: GB:X55664

C:Superfamily: phosphoenolpyruvate carboxylase

C:Keywords: carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase; phosphoprotein

Query Match 48.8%; Score 42; DB 2; Length 960;  
Best Local Similarity 47.6%; Pred. No. 68;  
Matches 10; Conservative 1; Mismatches 2; Indels 8; Gaps 1;

QY 4 LRSNYE-----VKGRDV 16  
||||| :|||  
Db 850 LRKNYEETKELLQVAGHKDV 870

## RESULT 8

S31159

Phosphoenolpyruvate carboxylase (EC 4.1.1.31) CP21 - sorghum

C:Species: Sorghum bicolor (sorghum)

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 22-Jun-1999

C:Accession: S31159

R:Lepiniec, L.; Keryer, E.; Philippe, H.; Gadai, P.; Cretin, C.

Plant Mol. Biol. 21, 487-502, 1993

A:Title: Sorghum phosphoenolpyruvate carboxylase gene family: structure, function and m

A:Reference number: S31159; MUID:93184205; PMID:8443342

A:Accession: S31159

A:Molecule type: DNA

A:Residues: 1-960 <LEP>

A:Cross-references: EMBL:X65137; NID:g22614; PIDN:CAA46267.1; PID:g22615

C:Genetics:

A:Introns: 52/3; 183/3; 211/3; 285/3; 320/3; 349/3; 401/3; 734/3; 863/3

C:Superfamily: phosphoenolpyruvate carboxylase

C:Keywords: carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase; phosphoprotein

Query Match 48.8%; Score 42; DB 2; Length 960;  
Best Local Similarity 47.6%; Pred. No. 68;  
Matches 10; Conservative 1; Mismatches 2; Indels 8; Gaps 1;

QY 4 LRSNYE-----VKGRDV 16  
||||| :|||  
Db 850 LRKNYEETKELLQVAGHKDV 870

## RESULT 9

QYNT

phosphoenolpyruvate carboxylase (EC 4.1.1.31) - common tobacco

C:Species: Nicotiana tabacum (common tobacco)

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 18-Jun-1999

C:Accession: S17440

R:Koizumi, N.; Sato, F.; Terano, Y.; Yamada, Y.

Plant Mol. Biol. 17, 535-539, 1991

A:Title: Sequence analysis of cDNA encoding phosphoenolpyruvate carboxylase from culture

A:Reference number: S17440; MUID:91355949; PMID:1884006

A:Accession: S17440

A:Molecule type: mRNA

A:Residues: 1-964 <KOI>

A:Cross-references: EMBL:X59016; NID:g22588; PIDN:CAA41758.1; PID:g22589

C:Superfamily: phosphoenolpyruvate carboxylase

C:Keywords: carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase; phosphoprotein

Query Match 48.8%; Score 42; DB 1; Length 964;  
Best Local Similarity 42.9%; Pred. No. 69;  
Matches 9; Conservative 3; Mismatches 1; Indels 8; Gaps 1;

QY 4 LRSNYE-----VKGRDV 16  
||||| :|||  
Db 854 LRSNYETRSLLQIAGHKDL 874

## RESULT 10

S25081

phosphoenolpyruvate carboxylase (EC 4.1.1.31) - Flaveria pringlei

## S18318

phosphoenolpyruvate carboxylase (EC 4.1.1.31) isoform C4 (clone ppcl-1) - Flaveria t

C:Species: Flaveria trinervia

C:Date: 22-Nov-1993 #sequence\_revision 23-Feb-1996 #text\_change 22-Jun-1999

C:Accession: S18318; S60517; S17342

R:Poetsch, W.; Hermans, J.; Westhoff, P.

FEBS Lett. 292, 133-136, 1991

A:Title: Multiple cDNAs of phosphoenolpyruvate carboxylase in the C(4) dicot Flaveri

A:Reference number: S18318; MUID:92070471; PMID:1720398

A:Accession: S18318

A:Molecule type: mRNA

A:Residues: 1-966 <POE>

A:Cross-references: EMBL:X61304

A:Note: only a part of the coding sequence is given in this paper

R:Poetsch, W.

submitted to the EMBL Data Library, June 1994

A:Reference number: S60517

A:Accession: S60517

A:Molecule type: mRNA

A:Residues: 1-256, 'R', 258-852, 'S', 854-966 <POF>

A:Cross-references: EMBL:X61304; NID:g498698; PIDN:CAA43601.1; PID:g498699

A:Note: this is a revision to the sequence from reference S17342

R:Poetsch, W.; Hermans, J.; Westhoff, P.

submitted to the EMBL Data Library, August 1991

A:Reference number: S17342

A:Accession: S17342

A:Molecule type: mRNA

A:Residues: 1-180, 'SMEGFVIWVPS', 191, 'MPKTSLLM', 200, 'SRNMSCFK', 211-852, 'S', 854-966 <F

A:Cross-references: EMBL:X61304

A:Note: this sequence has been revised in reference S60517

C:Superfamily: phosphoenolpyruvate carboxylase

C:Keywords: carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase; phosphoprote

Query Match 48.8%; Score 42; DB 2; Length 966;

Best Local Similarity 36.4%; Pred. No. 69;

Matches 8; Conservative 5; Mismatches 1; Indels 8; Gaps 1;

QY 3 FLRSNYE-----VKGRDV 16

||||| :|||

Db 853 YLRANYEETKYNLLKIAGHKDL 874

## RESULT 11

S52853

phosphoenolpyruvate carboxylase (EC 4.1.1.31) - Flaveria pringlei

C:Species: Flaveria pringlei

C:Date: 19-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 22-Jun-1999

C:Accession: S52853

R:Svensson, P.; Blaessing, O.E.; Westhoff, P.

submitted to the EMBL Data Library, April 1995

A:Description: Characteristics of the orthologous C4 and C3 PPCA phosphoenolpyruvate

A:Reference number: S52853

A:Accession: S52853

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-966 <SVE>

A:Cross-references: EMBL:Z48966; NID:g763096; PIDN:CAA88829.1; PID:g763097

C:Superfamily: phosphoenolpyruvate carboxylase

C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 48.8%; Score 42; DB 2; Length 966;

Best Local Similarity 42.9%; Pred. No. 69;

Matches 9; Conservative 3; Mismatches 1; Indels 8; Gaps 1;

QY 4 LRSNYE-----VKGRDV 16

||||| :|||

Db 854 LRANYEETKYLLKIAGHKDL 874

## RESULT 12

S25081

phosphoenolpyruvate carboxylase (EC 4.1.1.31) - Flaveria pringlei

C:Species: Flaveria pringlei  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 22-Jun-1999  
C:Accession: S25081  
R:Hermans, J.; Westhoff, P.  
Mol. Gen. Genet. 234, 275-284, 1992  
A:Title: Homologous genes for the C(4) isoform of phosphoenolpyruvate carboxylase in a C  
A:Reference number: S25081; MUID:92374996; PMID:1508152  
A:Accession: S25081  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-967 <HER>  
A:Cross-references: EMBL:X64144; NID:g18457; PIDN:CAA45505.1; PID:g18458  
C:Genetics:  
A:Gene: ppcA1  
A:Introns: 56/3; 187/2; 215/3; 291/2; 325/3; 355/1; 406/3; 739/3; 868/3  
C:Superfamily: phosphoenolpyruvate carboxylase  
C:Keywords: carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase; phosphoprotein  
Query Match 48.8%; Score 42; DB 2; Length 967;  
Best Local Similarity 42.9%; Pred. No. 69;  
Matches 9; Conservative 3; Mismatches 1; Indels 8; Gaps 1;  
QY 4 LRSNYE-----VKGHRDV 16  
|||:|||||:|||||:  
DB 855 LRANYEETKDYLLKTIAGHRDL 875  
RESULT 13  
B71972  
DNA topoisomerase I - Helicobacter pylori (strain J99)  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 29-Sep-1999  
C:Accession: B71972  
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
A:Reference number: A71800; MUID:99120557; PMID:9923382  
A:Accession: B71972  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-736 <ARN>  
A:Cross-references: GB:AE001450; GB:AE001439; NID:g4154617; PIDN:RAD05695.1; PID:g415462  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: topA\_1  
C:Superfamily: DNA topoisomerase I  
Query Match 47.7%; Score 41; DB 2; Length 736;  
Best Local Similarity 55.0%; Pred. No. 76;  
Matches 11; Conservative 2; Mismatches 3; Indels 4; Gaps 2;  
QY 1 RGFLRSNYEY---KGH-RDV 16  
: || ||||| |||||:  
DB 17 KNFLDKNYEVVASKGHVRDL 36  
RESULT 14  
D64534  
DNA topoisomerase I - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 29-Sep-1999  
C:Accession: D64534  
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;  
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen  
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.  
Nature 388, 539-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A:Reference number: A64520; MUID:97394467; PMID:9252185  
A:Accession: D64534  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
A:Residues: 1-736 <TOM>  
A:Cross-references: GB:AE000533; GB:AE000511; NID:g2313196; PIDN:RAD07184.1; PID:g231  
C:Superfamily: DNA topoisomerase I  
Query Match 47.7%; Score 41; DB 2; Length 736;  
Best Local Similarity 55.0%; Pred. No. 76;  
Matches 11; Conservative 2; Mismatches 3; Indels 4; Gaps 2;  
QY 1 RGFLRSNYEY---KGH-RDV 16  
: || ||||| |||||:  
DB 17 KNFLDKNYEVVASKGHVRDL 36  
RESULT 15  
T01058  
Hypothetical protein YUP8H12R.42 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 05-Feb-1999 #sequence\_revision 05-Feb-1999 #text\_change 22-Oct-1999  
C:Accession: T01058  
R:Theologis, A.; Vysotskaia, V.S.; Osborne, B.I.; Schwartz, J.R.; Federspiel, N.A.; K  
Oefner, P.; Davis, R.W.  
submitted to the EMBL Data Library, May 1998  
A:Description: Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence.  
A:Reference number: Z14227  
A:Accession: T01058  
A:Status: translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-830 <THE>  
A:Cross-references: EMBL:AC002986; NID:g2494106; PID:g3152574; PIDN:AAC17055.1; GSPDB  
C:Genetics:  
A:Gene: ATSP:YUP8H12R.42  
A:Map position: 1  
A:Introns: 68/3; 130/3; 160/3; 223/3; 444/3; 482/3; 522/3; 563/3; 596/3; 615/3; 679/2  
C:Superfamily: yeast lanosterol synthase  
Query Match 47.7%; Score 41; DB 2; Length 830;  
Best Local Similarity 50.0%; Pred. No. 86;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
QY 1 RGFLRSNYEVKGHVDV 16  
|||||:|||||:  
DB 48 RGFLDNFRVRKGCSDL 63  
Search completed: November 4, 2002, 20:37:55  
Job time: 52 secs

Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 10 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x MWRNATWS1 (1-678)

QY 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 466 CGGGCTTCTGAGATCAACTATGAGTCAAGACACCGGATGTG 513

RESULT 4  
 AF076482

LOCUS AF076482 680 bp mRNA linear ROD 15-AUG-1998  
 DEFINITION Mus musculus peptidoglycan recognition protein precursor (Pgrp)  
 mRNA, complete cds.

ACCESSION AF076482  
 VERSION AF076482.1 GI:3342530

KEYWORDS  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 680)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Kang,D., Liu,G., Lundstrom,A., Gellius,E. and Steiner,H.  
 TITLE A peptidoglycan recognition protein in innate immunity conserved  
 from insects to humans

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (17), 10078-10082 (1998)  
 MEDLINE 98374308  
 PUBMED 9707603

REFERENCE 2 (bases 1 to 680)

AUTHORS Kang,D., Liu,G., Lundstrom,A., Gellius,E. and Steiner,H.

TITLE Direct Submission  
 JOURNAL Submitted (08-JUL-1998) Microbiology, Stockholm University,  
 Stockholm S-106 91, Sweden

FEATURES  
 Location/Qualifiers

1..680  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /tissue\_type="spleen"

gene 1..680

CDS /gene="Pgrp"

36..584 /gene="Pgrp"

/note="PGRP"

/codon\_start=1

/product="peptidoglycan recognition protein precursor"

/protein\_id="AAC31821.1"

/db\_xref="GI:3342531"

/translation="MLFACALLALLGLATSCSFIVPRSEWRALPSCSSRLGHPVRYV  
 VISHTAGSFCNSPDCEQARNVQHYHKNELGCMQVAYNFLIGDGHVYEGRWNIKG  
 DHTGPIWNPMSIGITFMGNFMDRVPKRALRALNLECGVSRGFLRSNYEVKGHRDV  
 OSTLSPGDQLYQVQSWEHYRE"

sig\_peptide 36..83

/gene="Pgrp"

84..581 /gene="Pgrp"

/product="peptidoglycan recognition protein"

BASE COUNT 146 a 216 c 174 g 144 t

ORIGIN

Alignment Scores:  
 Pred. No.: 5.63e-07 Length: 680  
 Score: 86.00 Matches: 16  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 10 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x AF076482 (1-680)

QY 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16  
 ||||||||||||||||||||||||||||||||||||||||||||||||

Db 468 CGGGCTTCTGAGATCAACTATGAGTCAAGACACCGGATGTG 515

RESULT 5

BC005582

LOCUS BC005582

DEFINITION

ACCESSION BC005582

VERSION BC005582.1

KEYWORDS

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

NIH-MGC Project URL: http://mgc.nci.nih.gov

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

CDNA Library Preparation: Gilbert Smith, Ph.D.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: http://www.hgsc.bcm.tmc.edu/cdna/

Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,

Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,

Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Series: IRAC Plate: 17 Row: j Column: 22

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 6679292.

FEATURES

Location/Qualifiers

1..713

/organism="Mus musculus"

/db\_xref="taxon:10090"

/map="FVB/N"

/clone="MGC:11430 IMAGE:3969014"

/tissue\_type="Mammary tumor. Metallothionien-TGF alpha

model. 10 month old virgin mouse. Taken by biopsy."

/clone\_lib="NCI\_CGAP\_Maml"

/lab\_host="DH10B"

/notes="Vector: pCMV-SPORT6"

47..595

/codon\_start=1

/product="peptidoglycan recognition protein"

/protein\_id="AAH05582.1"

/db\_xref="GI:13542756"

/db\_xref="LocusID:21946"

/translation="MLFACALLALLGLATSCSFIVPRSEWRALPSCSSRLGHPVRYV

VISHTAGSFCNSPDCEQARNVQHYHKNELGCMQVAYNFLIGDGHVYEGRWNIKG

DHTGPIWNPMSIGITFMGNFMDRVPKRALRALNLECGVSRGFLRSNYEVKGHRDV

OSTLSPGDQLYQVQSWEHYRE"

CDS

BASE COUNT 172 a 218 c 178 g 145 t

ORIGIN

Alignment Scores:

Pred. No.: 5.93e-07 Length: 713

Score: 86.00 Matches: 16

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 10 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x BC005582 (1-713)

Qy 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16  
|||||  
Db 479 CGGGCTCTCTGAGATCAACTATGAGTCAAGGACACCGGATGTG 526

RESULT 6  
LOCUS AF154114 630 bp mRNA linear ROD 01-JUN-2000  
DEFINITION Rattus norvegicus peptidoglycan recognition protein PGRP (Pgrp)  
mRNA, complete cds.  
AF154114  
VERSION AF154114.1 GI:8132325  
KEYWORDS  
SOURCE Rattus norvegicus.  
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 630)  
Rehman, A., Teodecki, E. E. and Krueger, J. M.  
Direct Submission  
Submitted (25-MAY-1999) Veterinary Comparative Anatomy,  
Pharmacology, and Physiology, Washington State University, P.O. Box  
646520, Pullman, WA 99164, USA  
Location/Qualifiers  
1. .630  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/tissue\_type="spleen"  
1. .630  
/gene="Pgrp"  
20. 571  
/gene="Pgrp"

/note="similar to Homo sapiens and Mus musculus  
peptidoglycan recognition protein; similar to Trichoplusia  
ni and Mus musculus TAG7"  
/codon\_start=1  
/product="peptidoglycan recognition protein PGRP"  
/protein\_id="AF73252.1"  
/db\_xref="GI:8132326"

/translation="MLFAWPPALGLADSCFVVRSEWKLFPSECSKGLKPKVRY  
VVISHTAGSFCSPOSCEQARNVOLYQMKQWCDVAYNFLIGEDGHVYEGRGWTIK  
GDHTGPIWPMISGIFTMGDYSHRVPKRAALNLLKCGVSEGLRSEYEVKGRD  
VOSTLSPGDLVEIIGSDWBYRE"  
BASE COUNT 134 a 193 c 166 g 137 t  
ORIGIN

Alignment Scores:  
Pred. No.: 4.68e-06 Length: 630  
Score: 81.00 Matches: 15  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 94.19% Indels: 0  
DB: 10 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x AF154114 (1-630)  
Qy 2 GlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16  
|||||  
Db 458 GGCTCTGAGATCTAACTAGGTCAAGGACATCGAGATGTG 502

RESULT 7  
LOCUS AC110846 182897 bp DNA linear HTG 17-JUL-2002  
DEFINITION Rattus norvegicus clone CH230-105N7, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 77 unordered pieces.  
AC110846  
AC110846.3 GI:21744401  
VERSION HTG: HTGS.PHASE1.  
KEYWORDS Norway rat.  
SOURCE Rattus norvegicus  
ORGANISM

# REFERENCE AUTHORS

1 (bases 1 to 182897)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,  
Alsbrooks, S.L., Amarantunge, H.C., Are, J.R., Ayele, M., Banks, T.,  
Barberia, J., Benton, J., Bimager, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buha, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carroll, T., Carter, M., Cavazos, S.R., Chacko, J., Chaves, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
DeLaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,  
Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
Karlisson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Lounsegh, H.,  
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,  
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,  
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalobos, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wleczyk, R., Woodson, S., Worley, K.,  
Wu, C., Wu, Y., Wu, F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.

# REFERENCE AUTHORS

2 (bases 1 to 182897)  
Worley, K.C.  
Direct Submission  
Submitted (16-FEB-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 182897)  
Worley, K.C.  
Direct Submission  
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 13, 2002 this sequence version replaced gi:18767308.

# REFERENCE AUTHORS

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
Center project name: GRPZ  
Center clone name: CH230-105N7  
Summary Statistics  
Sequencing vector: Plasmid  
Assembly program: Phrap; version 0.990329  
Consensus quality: 107348 bases at least Q40  
Consensus quality: 113763 bases at least Q30

# REFERENCE AUTHORS

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
Center project name: GRPZ  
Center clone name: CH230-105N7  
Summary Statistics  
Sequencing vector: Plasmid  
Assembly program: Phrap; version 0.990329  
Consensus quality: 107348 bases at least Q40  
Consensus quality: 113763 bases at least Q30

# REFERENCE AUTHORS

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
Center project name: GRPZ  
Center clone name: CH230-105N7  
Summary Statistics  
Sequencing vector: Plasmid  
Assembly program: Phrap; version 0.990329  
Consensus quality: 107348 bases at least Q40  
Consensus quality: 113763 bases at least Q30

Consensus quality: 119211 bases at least Q20

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 77 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
1170: contig of 1170 bp in length  
1270: gap of unknown length  
1271: contig of 1013 bp in length  
2284: gap of unknown length  
2384: contig of 1024 bp in length  
3408: gap of unknown length  
3508: contig of 1056 bp in length  
4564: gap of unknown length  
4664: contig of 1039 bp in length  
5703: gap of unknown length  
5803: contig of 1227 bp in length  
7030: gap of unknown length  
7130: contig of 1175 bp in length  
8404: gap of unknown length  
8405: contig of 1057 bp in length  
9462: gap of unknown length  
9562: contig of 1409 bp in length  
10971: gap of unknown length  
11071: contig of 1658 bp in length  
12729: gap of unknown length  
12829: contig of 1558 bp in length  
14387: gap of unknown length  
14487: contig of 1158 bp in length  
15645: gap of unknown length  
15745: contig of 1464 bp in length  
17209: gap of unknown length  
17309: contig of 1301 bp in length  
18610: gap of unknown length  
18710: contig of 1893 bp in length  
20603: gap of unknown length  
20703: contig of 1402 bp in length  
22105: gap of unknown length  
22205: contig of 1453 bp in length  
23658: gap of unknown length  
23758: contig of 1890 bp in length  
25648: gap of unknown length  
25748: contig of 1064 bp in length  
26912: gap of unknown length  
26913: contig of 1866 bp in length  
28778: gap of unknown length  
28878: contig of 1135 bp in length  
30013: gap of unknown length  
30113: contig of 1362 bp in length  
31475: gap of unknown length  
31575: contig of 1651 bp in length  
33226: gap of unknown length  
33227: contig of 1329 bp in length  
34654: gap of unknown length  
34755: contig of 1045 bp in length  
35799: gap of unknown length  
35800: contig of 1339 bp in length  
37239: gap of unknown length  
37339: contig of 1956 bp in length  
39295: gap of unknown length  
39395: contig of 1336 bp in length  
40731: gap of unknown length  
40831: contig of 2285 bp in length  
43216: gap of unknown length  
43216: contig of 1492 bp in length  
44708: gap of unknown length  
44808: contig of 1526 bp in length

46334: gap of unknown length  
48307: contig of 1874 bp in length  
48407: gap of unknown length  
48408: contig of 1293 bp in length  
49800: gap of unknown length  
51925: contig of 2125 bp in length  
52025: gap of unknown length  
54497: contig of 2472 bp in length  
54598: gap of unknown length  
56275: contig of 1678 bp in length  
56375: gap of unknown length  
57816: contig of 1441 bp in length  
57817: gap of unknown length  
59765: contig of 1849 bp in length  
59865: gap of unknown length  
62577: contig of 2712 bp in length  
62677: gap of unknown length  
64206: contig of 1529 bp in length  
64306: gap of unknown length  
67434: contig of 3128 bp in length  
67534: gap of unknown length  
69462: contig of 1928 bp in length  
69562: gap of unknown length  
70978: contig of 1416 bp in length  
71078: gap of unknown length  
73252: contig of 2174 bp in length  
73352: gap of unknown length  
76275: contig of 2923 bp in length  
76375: gap of unknown length  
79483: contig of 3108 bp in length  
79584: gap of unknown length  
82304: contig of 2721 bp in length  
82404: gap of unknown length  
84815: contig of 2411 bp in length  
84915: gap of unknown length  
87439: contig of 2524 bp in length  
87539: gap of unknown length  
90118: contig of 2579 bp in length  
90218: gap of unknown length  
92751: contig of 2533 bp in length  
92851: gap of unknown length  
94485: contig of 1634 bp in length  
94585: gap of unknown length  
96642: contig of 2057 bp in length  
96742: gap of unknown length  
99345: contig of 2603 bp in length

# Alignment Scores:

Pred. No.: 0.00227 Length: 182897  
Score: 81.00 Matches: 15  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 94.19% Indels: 0  
DB: 2 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x AC110846 (1-182897)

QY 2 GlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16  
|||||  
Db 127455 GCCTCTGAGTCTACTATGAGGTCAGAGGACATCGAGATGTG 127499

## RESULT 8

CDRI31676  
LOCUS Camelus dromedarius mRNA for peptidoglycan recognition protein. 700 bp mRNA linear MAM 21-DEC-2000  
DEFINITION AJ131676  
ACCESSION AJ131676  
VERSION AJ131676.1 GI:11990123  
KEYWORDS peptidoglycan recognition protein.  
SOURCE Arabian camel.  
ORGANISM Camelus dromedarius  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.  
REFERENCE 1 (bases 1 to 700)







C:Species: Flaveria pringlei  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 22-Jun-1999  
C:Accession: S25081  
R:Hermans, J.; Westhoff, P.  
Mol. Gen. Genet. 234, 275-284, 1992  
A:Title: Homologous genes for the C(4) isoform of phosphoenolpyruvate carboxylase in a C  
A:Reference number: S25081; MUID:92374996; PMID:1508152  
A:Accession: S25081  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-967 <HER>  
A:Cross-references: EMBL:X64144; NID:g18457; PIDN:CRAA5505.1; PID:g18458  
C:Genetics:  
A:Gene: ppcA1  
A:Introns: 56/3; 187/2; 215/3; 291/2; 325/3; 355/1; 406/3; 739/3; 868/3  
C:Superfamily: phosphoenolpyruvate carboxylase  
C:Keywords: carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase; phosphoprotein

Query Match 48.8%; Score 42; DB 2; Length 967;  
Best Local Similarity 42.9%; Pred. No. 69;  
Matches 9; Conservative 3; Mismatches 1; Indels 8; Gaps 1;

QY 4 LRSNVE-----VKGHRDV 16  
II:III  
Db 855 LRANVEETKDYLLKXIAGHRDL 875

RESULT 13  
B71972  
DNA topoisomerase I - Helicobacter pylori (strain J99)  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 29-Sep-1999  
C:Accession: B71972  
R:Alm, R.A.; Ling, R.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
A:Reference number: A71800; MUID:99120557; PMID:99233682  
A:Accession: B71972  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-736 <ARN>  
A:Cross-references: GB:AE001450; GB:AE001439; NID:g4154617; PIDN:RAD05695.1; PID:g415462  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: topA\_1  
C:Superfamily: DNA topoisomerase I

Query Match 47.7%; Score 41; DB 2; Length 736;  
Best Local Similarity 55.0%; Pred. No. 76;  
Matches 11; Conservative 2; Mismatches 3; Indels 4; Gaps 2;

QY 1 RGFLRSNVEV---KGH-RDV 16  
: II IIII IIII  
Db 17 KNFLDKNYEVWASKGHVRDL 36

RESULT 14  
D64534  
DNA topoisomerase I - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 29-Sep-1999  
C:Accession: D64534  
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;  
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne  
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.  
Nature 388, 539-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A:Reference number: A64520; MUID:97394467; PMID:9252185  
A:Accession: D64534  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
A:Residues: 1-736 <TOM>  
A:Cross-references: GB:AE000533; GB:AE000511; NID:g2313196; PIDN:RAD07184.1; PID:g231  
C:Superfamily: DNA topoisomerase I

Query Match 47.7%; Score 41; DB 2; Length 736;  
Best Local Similarity 55.0%; Pred. No. 76;  
Matches 11; Conservative 2; Mismatches 3; Indels 4; Gaps 2;

QY 1 RGFLRSNVEV---KGH-RDV 16  
: II IIII IIII  
Db 17 KNFLDKNYEVIAASKGHVRDL 36

RESULT 15  
T01058  
hypothetical protein YUP8H12R.42 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 05-Feb-1999 #sequence\_revision 05-Feb-1999 #text\_change 22-Oct-1999  
C:Accession: T01058  
R:Theologis, A.; Vysotskaia, V.S.; Osborne, B.I.; Schwartz, J.R.; Federspiel, N.A.;  
Oefner, P.; Davis, R.W.  
submitted to the EMBL Data Library, May 1998  
A:Description: Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence.  
A:Reference number: Z14227  
A:Accession: T01058  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-830 <THE>  
A:Cross-references: EMBL:AC002986; NID:g2494106; PID:g3152574; PIDN:AAC17055.1; GSPD  
C:Genetics:  
A:Gene: ATSP:YUP8H12R.42  
A:Map position: 1  
A:Introns: 68/3; 130/3; 160/3; 223/3; 444/3; 482/3; 522/3; 563/3; 596/3; 615/3; 679/  
C:Superfamily: yeast lanosterol synthase

Query Match 47.7%; Score 41; DB 2; Length 830;  
Best Local Similarity 50.0%; Pred. No. 86;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 RGFLRSNVEVKGHRDV 16  
IIII: : IIII  
Db 48 RGFLDNFRVRKGCSDL 63

Search completed: November 4, 2002, 20:37:55  
Job time : 52 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 4, 2002, 19:07:41 ; Search time 36 Seconds  
(without alignments)  
18.434 Million cell updates/sec

Title: US-09-462-625-2\_copy\_145\_160  
Perfect score: 86  
Sequence: 1 RGFRLSNYEVKGRDV 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	86	100.0	182	1	PGRP_MOUSE
2	64	74.4	196	1	PGRP_HUMAN
3	48	55.8	182	1	PGRP_TRINI
4	43	50.0	960	1	CAP2_SORBI
5	43	50.0	967	1	CAP2_MAIZE
6	42	48.8	960	1	CAP1_SORBI
7	42	48.8	964	1	CAPP_TOBAC
8	42	48.8	967	1	CAP1_FLAPR
9	41	47.7	123	1	ANG2_BOVIN
10	41	47.7	447	1	VIPR_CARAU
11	41	47.7	736	1	TOP1_HELPJ
12	41	47.7	736	1	TOP1_HELPY
13	41	47.7	1065	1	KDGI_HUMAN
14	40	46.5	806	1	LON_BORBU
15	40	46.5	809	1	PUPB_PSEPU
16	40	46.5	819	1	LON_CHLMU
17	40	46.5	819	1	LON_CHLPM
18	40	46.5	819	1	LON_CHLTR
19	40	46.5	966	1	CAP1_MESCR
20	40	46.5	966	1	CAP1_SACHY
21	40	46.5	1025	1	THPE_RATY
22	39	45.3	651	1	Y942_METJA
23	39	45.3	813	1	LON2_BORBU
24	39	45.3	966	1	CAP2_FLATR
25	39	45.3	966	1	CAPP_FLAAU
26	39	45.3	967	1	CAP1_FLATR
27	38	44.2	264	1	IOD2_RANCA
28	38	44.2	370	1	YB11_HALNI
29	38	44.2	1024	1	THDE_HUMAN
30	37.5	43.6	340	1	PURR_SALTY
31	37	43.0	246	1	PURR_METJA
32	37	43.0	293	1	YKUM_BACSU
33	37	43.0	332	1	HEM2_RHOCA

## RESULT 1

ID	PGRP_MOUSE	STANDARD	PRT	182 AA
AC	O88593; O62185;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Peptidoglycan recognition protein precursor (Cytokine tag7).			
GN	PGLYRP OR PGRP OR TAG7.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.			
RC	TISSUE-Spleen;			
RX	MEDLINE=98374308; PubMed=9707603;			
RA	Kang D., Liu G., Lundstroem A., Gellus E., Steiner H.;			
RT	"A peptidoglycan recognition protein in innate immunity conserved from insects to humans."			
RL	Proc. Natl. Acad. Sci. U.S.A. 95:10078-10082(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.			
RX	MEDLINE=98325081; PubMed=9660837;			
RA	Kiselev S.L., Kustikova O.S., Korobko E.V., Prokhortchouk E.B.,			
RA	Kabishhev A.A., Lukandin E.M., Georgiev G.P.;			
RT	"Molecular cloning and characterization of the mouse tag7 gene encoding a novel cytokine."			
RL	J. Biol. Chem. 273:18633-18639(1998).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Slayton W.B., Rigau A., Hancock J.D., Zaugg J.K., Le T.V.,			
RA	Trautman M.S., Spangrude G.J., Carroll W.L., Schibler K.R.;			
RT	"Granulocyte-colony stimulating factor up-regulates expression of murine tag7 during myeloid differentiation."			
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Small intestine;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,			
RA	Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,			

## ALIGNMENTS

34	37	43.0	337	1	INT2_RCOLI
35	37	43.0	357	1	VNS2_BTVIS
36	37	43.0	406	1	REP_CLOPE
37	37	43.0	473	1	STYM_SCHPO
38	37	43.0	505	1	CHOL_HUMAN
39	37	43.0	513	1	LEU1_IACLA
40	37	43.0	521	1	SYK_BORBU
41	37	43.0	610	1	VEL_HPV60
42	37	43.0	777	1	LON_BUCAL
43	37	43.0	795	1	LON_AQUAE
44	37	43.0	810	1	LON_AZOBR
45	37	43.0	887	1	CYA2_HUMAN

P09999 escherichia  
P32932 clostridium  
P18016 clostridium  
O13969 schizosacch  
Q9V236 homo sapien  
Q02141 lactococcus  
O51603 borrelia bu  
Q80943 human papil  
P57549 buchmera ap  
O66605 aquifex aeo  
P77810 azospirillu  
Q08462 homo sapien

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.,  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: BINDS SPECIFICALLY TO PEPTIDOGLYCAN AND IS INVOLVED IN  
CC INNATE IMMUNITY. THE SOLUBLE FORM TRIGGERS APOPTOSIS IN VITRO.  
CC -1- SUBCELLULAR LOCATION: EXISTS IN BOTH SOLUBLE AND MEMBRANE-  
CC ASSOCIATED FORMS.  
CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN SPLEEN AND LUNG. ALSO  
CC DETECTED IN BRAIN AND THYMUS. IN THE LUNG, EXPRESSED IN THE  
CC INTRAALVEOLAR SPACE, IN THE BRAIN, EXPRESSED IN THE PURKINJE CELLS  
CC OF THE CEREBELLUM AND IN CERTAIN LAYERS OF NEURONS IN THE  
CC HIPPOCAMPUS. ALSO DETECTED IN CELLS FILLING THE SPACE WITHIN THE  
CC INTESTINAL VILLOS.  
CC -1- SIMILARITY: BELONGS TO THE PEPTIDOGLYCAN RECOGNITION PROTEIN  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF076482; AAC31021.1; -  
DR EMBL; AF193843; AAF06335.1; -  
DR EMBL; AK008335; BAB25611.1; -  
DR EMBL; BC005582; AAH05582.1; -  
DR EMBL; X86374; CAA60133.1; ALT\_SEQ.  
DR EMBL; Y12088; CAA72803.1; -  
DR MGI; MGI:1345092; Pglyrp.  
KW Immune response; Cytokine; Apoptosis; Signal.  
FT SIGNAL 1 18 POTENTIAL.  
FT CHAIN 19 182 PEPTIDOGLYCAN RECOGNITION PROTEIN.  
FT DISULFID 54 60 POTENTIAL.  
FT SEQUENCE 182 AA; 20489 MW; 9844E2137F047F14 CRC64;  
Query Match 100.0%; Score 86; DB 1; Length 182;  
Best Local Similarity 100.0%; Pred. No. 4.1e-08;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGFRLSNYEVKGHRDV 16  
DB 145 RGFRLSNYEVKGHRDV 160  
RESULT 2  
ID PGRP\_HUMAN STANDARD; PRT; 196 AA.  
AC 075594;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Peptidoglycan recognition protein precursor (SBB168).  
GN PGLYRP OR PGRP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.  
RP TISSUE=Bone marrow;  
RC MEDLINE=98374308; PubMed=9707603;  
RA Kang D., Liu G., Lundstrom A., Gelius E., Steiner H.;  
RT "A peptidoglycan recognition protein in innate immunity conserved from  
insects to humans."  
RL Proc. Natl. Acad. Sci. U.S.A. 95:10078-10082(1998).  
RN [2]

RP SEQUENCE FROM N.A.  
RA Lamerdin J.E., McCready P.M., Richardson P., Sakaladasis G.,  
RA Burkhardt-Schulz K., Gordon L., Scott D., Johnson G., Stillwagen S.,  
RA Phan H., Velasco N., Do L., Regala W., Terry A., Danganan L.,  
RA Erler A., Christensen M., Georgescu A., Avila J., Attix C.,  
RA Andreise T., Amico-Keller G., Coefield J., Duarte S., Lucas S.,  
RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A., Sanders C.,  
RA Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Wan T., Zhang W., Cao X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: BINDS SPECIFICALLY TO PEPTIDOGLYCAN AND IS INVOLVED IN  
CC INNATE IMMUNITY.  
CC -1- FUNCTION: THE SOLUBLE FORM TRIGGERS APOPTOSIS IN VITRO (BY  
CC SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: EXISTS IN BOTH SOLUBLE AND MEMBRANE-  
CC ASSOCIATED FORMS (BY SIMILARITY).  
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BONE MARROW. WEAK  
CC EXPRESSION FOUND IN KIDNEY, LIVER, SMALL INTESTINE, SPLEEN,  
CC THYMUS, PERIPHERAL LEUKOCYTE, LUNG AND FETAL SPLEEN.  
CC -1- SIMILARITY: BELONGS TO THE PEPTIDOGLYCAN RECOGNITION PROTEIN  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF076483; AAC31822.1; -  
DR EMBL; AC007785; AAD38243.1; -  
DR EMBL; AF242517; AAF99598.1; -  
DR HSP; P00806; IARO.  
DR Genew; HGNC:8904; PGLYRP.  
DR MIM; 604963; -  
KW Immune response; Cytokine; Apoptosis; Signal.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 196 PEPTIDOGLYCAN RECOGNITION PROTEIN.  
FT DISULFID 67 73 POTENTIAL.  
FT CARBOHYD 112 112 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT SEQUENCE 196 AA; 21731 MW; D954C51440DC27DC CRC64;  
Query Match 74.4%; Score 64; DB 1; Length 196;  
Best Local Similarity 75.0%; Pred. No. 0.00036;  
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 RGFRLSNYEVKGHRDV 16  
DB 158 QGALRSNYVLKGRDV 173  
RESULT 3  
ID PGRP\_TRINI STANDARD; PRT; 182 AA.  
AC 076537;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Peptidoglycan recognition protein precursor.  
GN PGRP.  
OS Trichoplusia ni (Cabbage looper).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;  
OC Ditrysia; Noctuoidea; Noctuidae; Plusiinae; Trichoplusia.  
OX NCBI\_TaxID=7111;  
[1]  
RN SEQUENCE FROM N.A., SEQUENCE OF 17-21, TISSUE SPECIFICITY, AND  
RP INDUCTION.  
RP TISSUE=Larva;

RX MEDLINE=98374308; PubMed=9707603;  
 RA Kang D., Liu G., Lundstroem A., Gelius E., Steiner H.;  
 RT "A peptidoglycan recognition protein in innate immunity conserved from  
 RL insects to humans.";  
 CC Proc. Natl. Acad. Sci. U.S.A. 95:10078-10082(1998).  
 CC -|- FUNCTION: BINDS SPECIFICALLY TO PEPTIDOGLYCAN AND TRIGGERS THE  
 CC PROPOXIDASE CASCADE WHICH IS AN IMPORTANT INSECT INNATE IMMUNE  
 CC DEFENSE MECHANISM.  
 CC -|- SUBUNIT: MONOMER (PROBABLE).  
 CC -|- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN FAT BODY WITH WEAK  
 CC EXPRESSION OBSERVED IN HEMOCYTE. NO EXPRESSION DETECTED IN  
 CC GUT.  
 CC -|- INDUCTION: BY BACTERIAL CHALLENGE.  
 CC -|- SIMILARITY: BELONGS TO THE PEPTIDOGLYCAN RECOGNITION PROTEIN  
 CC FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; AF076481; AAC31820.1; -  
 CC KW Immune response; Signal.  
 CC FT SIGNAL 1 16  
 CC FT CHAIN 17 182 PEPTIDOGLYCAN RECOGNITION PROTEIN.  
 CC FT DISULFID 18 140 POTENTIAL.  
 CC FT DISULFID 54 60  
 CC SQ SEQUENCE 182 AA; 20572 MW; 56631E762AE34794 CRC64;  
 CC  
 CC Query Match 55.8%; Score 48; DB 1; Length 182;  
 CC Best Local Similarity 50.0%; Pred. No. 0.23;  
 CC Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 CC  
 CC QY 1 RGFLRSNYEVKGRDGV 16  
 CC |||||  
 CC DB 144 RGLTANYHIVGHRQL 159  
 CC  
 CC RESULT 4  
 CC CAP2\_SORBI STANDARD; PRT; 960 AA.  
 CC AC P29194;  
 CC DT 01-DEC-1992 (Rel. 24, Created)  
 CC DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 CC DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 CC DE Phosphoenolpyruvate carboxylase 2 (EC 4.1.1.31) (PEPCASE) (CP28).  
 CC OS Sorghum bicolor (Sorghum) (Sorghum vulgare).  
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 CC OC Panicoideae; Andropogoneae; Sorghum.  
 CC OX NCBI\_TaxID=4558;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=92032766; PubMed=1840686;  
 CC RA Lepiniec L., Santl S., Kerker E., Amiet V., Vidal J., Gadal P.,  
 CC RA Cretin C.;  
 CC RT "Complete nucleotide sequence of one member of the Sorghum  
 CC plant Mol. Biol. 17:1077-1079(1991).  
 CC RL -|- FUNCTION: TO FORM OXALOACETATE, A FOUR-CARBON DICARBOXYLIC ACID  
 CC SOURCE FOR THE TRICARBOXYLIC ACID CYCLE.  
 CC CC -|- CATALYTIC ACTIVITY: Phosphate + oxaloacetate = H(2)O +  
 CC phosphoenolpyruvate + CO(2).  
 CC CC -|- PATHWAY: TRICARBOXYLIC ACID CYCLE. THIS ISOZYME IS INVOLVED IN  
 CC C3 PHOTOSYNTHESIS.  
 CC CC -|- SUBUNIT: HOMOTETRAMER.  
 CC CC -|- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC CC -|- SIMILARITY: BELONGS TO THE PEPCASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; X59925; CAA42549.1; -  
 CC DR PIR; S18240; S18240.  
 CC DR HSP; P00864; LFII.  
 CC DR InterPro; IPR001449; PEPCase.  
 CC DR Pfam; PF00311; PEPCase; 1.  
 CC DR PRINTS; PR00150; PEPCARBLASE.  
 CC DR PROSITE; PS00393; PEPCASE\_2; 1.  
 CC DR PROSITE; PS00781; PEPCASE\_1; 1.  
 CC KW Lyase; Carbon dioxide fixation; Allosteric enzyme; Multigene family;  
 CC KW Tricarboxylic acid cycle; Photosynthesis.  
 CC FT ACT\_SITE 167 BY SIMILARITY.  
 CC FT ACT\_SITE 167 BY SIMILARITY.  
 CC FT ACT\_SITE 595 BY SIMILARITY.  
 CC SQ SEQUENCE 960 AA; 109547 MW; 65FDC9B71156B18 CRC64;  
 CC  
 CC Query Match 50.0%; Score 43; DB 1; Length 960;  
 CC Best Local Similarity 47.6%; Pred. No. 12;  
 CC Matches 10; Conservative 2; Mismatches 1; Indels 8; Gaps 1;  
 CC  
 CC QY 4 LRSNYE-----VKGRDGV 16  
 CC |||||  
 CC DB 849 LRANYEETQKLLQVAGHRDL 869  
 CC  
 CC RESULT 5  
 CC CAP2\_MAIZE STANDARD; PRT; 967 AA.  
 CC ID CAP2\_MAIZE  
 CC AC P51059;  
 CC DT 01-OCT-1996 (Rel. 34, Created)  
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 CC DE Phosphoenolpyruvate carboxylase 2 (EC 4.1.1.31) (PEPCASE).  
 CC GN PEP4 OR PEP  
 CC OS Zea mays (Maize).  
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 CC OC Panicoideae; Andropogoneae; Zea.  
 CC OX NCBI\_TaxID=4577;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX STRAIN=CV. H84; TISSUE=Root;  
 CC RX MEDLINE=93054411; PubMed=1429504;  
 CC RA Kawamura T., Shigesada K., Toh H., Okumura S., Yanagisawa S.,  
 CC RA Izui K.;  
 CC RT "Molecular evolution of phosphoenolpyruvate carboxylase for C4  
 CC photosynthesis in maize: comparison of its cDNA sequence with a newly  
 CC isolated cDNA encoding an isozyme involved in the anaplerotic  
 CC function.";  
 CC RT J. Biochem. 112:147-154(1992).  
 CC RL -|- FUNCTION: TO FORM OXALOACETATE, A FOUR-CARBON DICARBOXYLIC ACID  
 CC SOURCE FOR THE TRICARBOXYLIC ACID CYCLE.  
 CC CC -|- CATALYTIC ACTIVITY: Phosphate + oxaloacetate = H(2)O +  
 CC phosphoenolpyruvate + CO(2).  
 CC CC -|- ENZYME REGULATION: BY LIGHT-REVERSIBLE PHOSPHORYLATION (BY  
 CC SIMILARITY).  
 CC CC -|- PATHWAY: TRICARBOXYLIC ACID CYCLE. THIS ISOZYME IS INVOLVED IN  
 CC C3 PHOTOSYNTHESIS.  
 CC CC -|- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC CC -|- SIMILARITY: BELONGS TO THE PEPCASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X61489; CAA43709.1; -
DR HSP; P00864; IFIY.
DR MaizeDB; 30066; -
DR InterPro; IPR001449; PEPCase.
DR Pfam; PF00311; PEPCase.1.
DR PRINTS; PR00150; PEPCARXLA.
DR PROSITE; PS00393; PEPCASE_2; 1.
DR PROSITE; PS00781; PEPCASE_1; 1.
DR Tricarboxylic acid cycle; Phosphorylation (BY SIMILARITY).
KW Tricarboxylic acid cycle; Phosphorylation (BY SIMILARITY).
FT MOD_RES 13 174
FT ACT_SITE 174 602 BY SIMILARITY.
FT ACT_SITE 602 602 BY SIMILARITY.
SQ SEQUENCE 967 AA; 10998 MW; 7034A2AD5521645B CRC64;

Query Match 50.08; Score 43; DB 1; Length 967;
Best Local Similarity 47.68; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 1; Indels 8; Gaps 1;

Qy 4 LRSNYE-----VKGRDV 16
   ||||| |
Db 856 LRANYEETKLLQVAGHRDL 876

RESULT 6
CAP1_SORBI STANDARD; PRT; 960 AA.
AC P29195;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Phosphoenolpyruvate carboxylase 1 (EC 4.1.1.31) (PEPCASE) (CP21).
GN PEPC.
OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Sorghum.
OX NCBI_TaxID=4558;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Root;
RX MEDLINE=91216449; PubMed=2022326;
RA Cretin C., Santi S., Keryer E., Lepiniec L., Tagu D., Vidal J.,
RA Gadal P.;
RT "The phosphoenolpyruvate carboxylase gene family of Sorghum: promoter
RT structures, amino acid sequences and expression of genes.";
RL Gene 99:87-94(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93184205; PubMed=8443342;
RA Lepiniec L., Keryer E., Philippe H., Gadal P., Cretin C.;
RT "Sorghum phosphoenolpyruvate carboxylase gene family: structure,
RT function and molecular evolution.";
RL Plant Mol. Biol. 21:487-502(1993).
CC -1- FUNCTION: TO FORM OXALOACETATE, A FOUR-CARBON DICARBOXYLIC ACID
CC -1- CATALYTIC ACTIVITY: Phosphate + oxaloacetate = H(2)O +
CC phosphoenolpyruvate + CO(2).
CC -1- ENZYME REGULATION: BY LIGHT-REVERSIBLE PHOSPHORYLATION (BY
CC SIMILARITY).
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE. THIS ISOZYME IS INVOLVED IN
CC C4 PHOTOSYNTHESIS.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE PEPCASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59016; CAA41758.1; -
DR PIR; S17440; QYNT.

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X55664; CAA39197.1; -
DR EMBL; X65137; CAA46267.1; -
DR PIR; JH0381; JH0381.
DR PIR; S31159; S31159.
DR HSP; P00864; IFIY.
DR InterPro; IPR001449; PEPCase.
DR Pfam; PF00311; PEPCase.1.
DR PRINTS; PR00150; PEPCARXLA.
DR PROSITE; PS00393; PEPCASE_2; 1.
DR PROSITE; PS00781; PEPCASE_1; 1.
DR Tricarboxylic acid cycle; Phosphorylation (BY SIMILARITY).
KW Tricarboxylic acid cycle; Phosphorylation (BY SIMILARITY).
FT MOD_RES 7 168
FT ACT_SITE 168 596 BY SIMILARITY.
FT ACT_SITE 596 596 BY SIMILARITY.
SQ SEQUENCE 960 AA; 109438 MW; 73927283CE3298AA CRC64;

Query Match 48.88; Score 42; DB 1; Length 960;
Best Local Similarity 47.68; Pred. No. 17;
Matches 10; Conservative 1; Mismatches 2; Indels 8; Gaps 1;

Qy 4 LRSNYE-----VKGRDV 16
   ||||| |
Db 850 LRKNEYETKLLQVAGHKDV 870

RESULT 7
CAP1_TOBAC STANDARD; PRT; 964 AA.
AC P27154;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Phosphoenolpyruvate carboxylase (EC 4.1.1.31) (PEPCASE).
GN PPC.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Samsun NN;
RX MEDLINE=91355949; PubMed=1884006;
RA Koizumi N., Sato F., Terano Y., Yamada Y.;
RT "Sequence analysis of cDNA encoding phosphoenolpyruvate carboxylase
RT from cultured tobacco cells.";
RL Plant Mol. Biol. 17:535-539(1991).
CC -1- FUNCTION: TO FORM OXALOACETATE, A FOUR-CARBON DICARBOXYLIC ACID
CC SOURCE FOR THE TRICARBOXYLIC ACID CYCLE.
CC -1- CATALYTIC ACTIVITY: Phosphate + oxaloacetate = H(2)O +
CC phosphoenolpyruvate + CO(2).
CC -1- ENZYME REGULATION: BY LIGHT-REVERSIBLE PHOSPHORYLATION (BY
CC SIMILARITY).
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE. THIS ISOZYME IS INVOLVED IN
CC C4 PHOTOSYNTHESIS.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE PEPCASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59016; CAA41758.1; -
DR PIR; S17440; QYNT.
```

DR pfam: PF00311; PEPcase; 1.  
DR PRINTS: PRO0150; PEPCaseX2.  
DR PROSITE: PS00393; PEPCaseL2; 1.  
DR PROSITE: PS00781; PEPCaseL1; 1.  
KW Lyase; Carbon dioxide fixation; Allosteric enzyme; Multigene family;  
KW Tricarboxylic acid cycle; Phosphorylation; Photosynthesis.  
FT MOD\_RES 11 11 PHOSPHORYLATION (BY SIMILARITY).  
FT ACT\_SITE 172 172 BY SIMILARITY.  
FT ACT\_SITE 601 601 BY SIMILARITY.  
FT CONFLICT 291 292 KH -> N (IN REF. 2).  
SQ SEQUENCE 967 AA; 110628 MW; 8317EA8078C038C0 CRC64;

Query Match 48.8%; Score 42; DB 1; Length 967;  
Best Local Similarity 42.9%; Pred. No. 18;  
Matches 9; Conservative 3; Mismatches 1; Indels 8; Gaps:

QY 4 LRSNYE-----VKGRDV 16  
||||| : |||||  
Db 855 LRANVEETKDYLLKIAGHRDL 875

RESULT 9  
ANG2\_BOVIN STANDARD; PRT; 123 AA.  
AC AC ANG2\_BOVIN STANDARD; PRT; 123 AA.  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Angiogenin-2 (EC 3.1.27.-).  
DS ANG2.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE.

DR TISSUP=Serum, and Milk;  
DR MEDLINE=97409980; PubMed=9266695;  
DR Straydon D.J., Bond M.D., Vallee B.L.;  
RA "An angiogenic protein from bovine serum and milk -- purification and  
RT primary structure of angiogenin-2.";  
RL Eur. J. Biochem. 247:535-544(1997).  
CC -1- FUNCTION: BINDS TIGHTLY TO PLACENTAL RIBONUCLEASE INHIBITOR AND  
CC HAS VERY LOW RIBONUCLEASE ACTIVITY. HAS POTENT ANGIOGENIC  
CC ACTIVITY. ANGIOGENIN INDUCES VASCULARIZATION OF NORMAL AND  
CC MALIGNANT TISSUES. ABOLISHES PROTEIN SYNTHESIS BY SPECIFICALLY  
CC HYDROLYZING CELLULAR TRNAS.  
CC -1- TISSUE SPECIFICITY: SERUM, AND MILK.  
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
DR HSSP: P10152; IAGI.  
DR InterPro: IPR001427; RNaseA.  
DR pfam: PF00074; RNaseA; 1.  
DR PRINTS: PD00794; RIBONUCLEASE.  
DR ProDom; PD000535; RNaseA; 1.  
DR SMART; SM00092; RNase\_Pc; 1.  
DR PROSITE: PS00127; RNASE\_PANCREATIC; 1.  
KW Hydrolase; Nuclease; Endonuclease; Angiogenesis;  
KW Protein synthesis inhibitor; Glycoprotein.  
MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
FT ACT\_SITE 12 12 BY SIMILARITY.  
FT ACT\_SITE 39 39 BY SIMILARITY.  
FT ACT\_SITE 113 113 BY SIMILARITY.  
FT DISULFID 25 80 BY SIMILARITY.  
FT DISULFID 38 91  
FT DISULFID 38 91  
FT DISULFID 56 106  
FT CARBOHYD 33 33  
SQ SEQUENCE 123 AA; 14522 MW; B703B9839919FD2F CRC64;

Query Match 47.7%; Score 41; DB 1; Length 123;  
Best Local Similarity 52.9%; Pred. No. 2.6;  
Matches 9; Conservative 2; Mismatches 4; Indels 2; Gaps:



FT ZN\_FING 663 689 C4-TYPE 3.  
 FT ZN\_FING 702 725 C4-TYPE 4.  
 FT ACT\_SITE 297 297 DNA\_CLEAVAGE (BY SIMILARITY).  
 SQ SEQUENCE 736 AA; 83262 MW; 779EF5655F3C14CF CRC64;

Query Match 47.7%; Score 41; DB 1; Length 736;  
 Best Local Similarity 55.0%; Pred. No. 19;  
 Matches 11; Conservative 2; Mismatches 3; Indels 4; Gaps 2;

QY 1 RGFSLRSNEYV---KGH-RDV 16  
 : || |||| |||| ||||  
 DB 17 KNFLDKNEYEVASKGHVRDL 36

## RESULT 12

ID TOPI\_HELPY STANDARD; PRT; 736 AA.  
 AC P55991;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA topoisomerase I (EC 5.99.1.2) (Omega-protein) (Relaxing enzyme)  
 DE (Unwinding enzyme) (Swivelase).  
 GN TOPA OR HP0116.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Helicobacter.  
 OX NCBI\_TaxID=210;  
 RN [1]

SEQUENCE FROM N.A.  
 RC STRAIN=26695 / ATCC 700392;  
 RX MEDLINE=97394467; PubMed=9252185;  
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,  
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,  
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
 RA Venter J.C.;

"The complete genome sequence of the gastric pathogen Helicobacter pylori";

RL Nature 388:539-547(1997).  
 CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE  
 CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.

CC -1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded  
 CC DNA, followed by passage and rejoining.

CC -1- SUBUNIT: MONOMER (BY SIMILARITY).

CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA  
 CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN

CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS  
 CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.

CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE  
 CC FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE000533; AAD07184.1;

CC HSSP: P06612; 1ECL.  
 CC TIGR: HP0116;

CC InterPro: IPR002936; DNAPrim\_toprim.  
 CC InterPro: IPR003601; DNATopI\_ATP\_bind.

CC InterPro: IPR003602; DNATopI\_DNA\_bind.

CC InterPro: IPR000380; Prok\_topisomase.

CC Pfam: PF01131; Topoisom\_bac; 1.

CC Pfam: PF01396; zf-C4\_Topoisom; 3.

DR Pfam: PF01751; Toprim; 1.  
 DR PRINTS: PRO0417; PRTPISMRASE1.  
 DR SMART: SM00437; TOP1AC; 1.  
 DR SMART: SM00436; TOP1BC; 1.  
 DR SMART: SM00493; TOP1RM; 1.  
 DR TIGRFAMs: TIGR01051; topa\_bact; 1.  
 DR PROSITE: PS00396; TOPOISOMERASE\_I\_PROK; 1.  
 KW Isomerase; Topoisomerase; DNA-binding; Zinc-finger; Metal-binding;  
 Repeat; Complete proteome.  
 FT ZN\_FING 572 598 C4-TYPE 1.  
 FT ZN\_FING 616 642 C4-TYPE 2.  
 FT ZN\_FING 663 689 C4-TYPE 3.  
 FT ZN\_FING 702 725 C4-TYPE 4.  
 FT ACT\_SITE 297 297 DNA\_CLEAVAGE (BY SIMILARITY).  
 SQ SEQUENCE 736 AA; 83195 MW; 357CE39735F48475 CRC64;

## Query Match

Best Local Similarity 47.7%; Score 41; DB 1; Length 736;  
 Matches 11; Conservative 2; Mismatches 3; Indels 4; Gaps 2;

QY 1 RGFSLRSNEYV---KGH-RDV 16  
 : || |||| |||| ||||  
 DB 17 KNFLDKNEYEVASKGHVRDL 36

## RESULT 13

ID KDGI\_HUMAN STANDARD; PRT; 1065 AA.  
 AC 075912; Q9NZ49;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Diacylglycerol kinase, iota (EC 2.7.1.107) (Diglyceride kinase) (DGK-  
 DE iota) (DAG kinase iota).  
 GN DGKI.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]

## SEQUENCE FROM N.A.

RP TISSUE=Retina;  
 RC MEDLINE=98047655; PubMed=9830018;  
 RA Ding L., Traer E., McIntyre T.M., Zimmerman G.A., Prescott S.M.;  
 RT "The cloning and characterization of a novel human diacylglycerol  
 RT kinase, DGK-iota.";  
 RL J. Biol. Chem. 273:32746-32752(1998).  
 RN [2]

SEQUENCE OF 135-1065 FROM N.A., AND VARIANT PHE-153.  
 RP MEDLINE=20173854; PubMed=10706894;

RA Bowne S.J., Sullivan L.S., Ding L., Traer E., Prescott S.M.,  
 RA Birch D.G., Kennan A., Humphries P., Daiger S.P.;

RT "Evaluation of human diacylglycerol kinase iota, DGKI, a homolog of  
 RT Drosophila rdgA, in inherited retinopathy mapping to 7q.";

RL Mol. Vision 6:6-9(2000).  
 CC -1- CATALYTIC ACTIVITY: ATP + 1,2-diacylglycerol = ADP + 1,2-  
 CC diacylglycerol 3-phosphate.

CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.

CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE  
 CC FAMILY.

CC -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG  
 CC BINDING DOMAINS.

CC -1- SIMILARITY: CONTAINS 2 ANK REPEATS.

CC -1- SIMILARITY: CONTAINS 1 MARCKS HOMOLOGY REGION.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).



```
DR EMBL; AF061936; AAC62010.1; .
DR EMBL; AF219939; AAF43006.1; .
DR EMBL; AF219907; AAF43006.1; JOINED.
DR EMBL; AF219908; AAF43006.1; JOINED.
DR EMBL; AF219909; AAF43006.1; JOINED.
DR EMBL; AF219910; AAF43006.1; JOINED.
DR EMBL; AF219911; AAF43006.1; JOINED.
DR EMBL; AF219912; AAF43006.1; JOINED.
DR EMBL; AF219913; AAF43006.1; JOINED.
DR EMBL; AF219914; AAF43006.1; JOINED.
DR EMBL; AF219915; AAF43006.1; JOINED.
DR EMBL; AF219916; AAF43006.1; JOINED.
DR EMBL; AF219917; AAF43006.1; JOINED.
DR EMBL; AF219918; AAF43006.1; JOINED.
DR EMBL; AF219919; AAF43006.1; JOINED.
DR EMBL; AF219920; AAF43006.1; JOINED.
DR EMBL; AF219921; AAF43006.1; JOINED.
DR EMBL; AF219922; AAF43006.1; JOINED.
DR EMBL; AF219923; AAF43006.1; JOINED.
DR EMBL; AF219924; AAF43006.1; JOINED.
DR EMBL; AF219925; AAF43006.1; JOINED.
DR EMBL; AF219926; AAF43006.1; JOINED.
DR EMBL; AF219927; AAF43006.1; JOINED.
DR EMBL; AF219928; AAF43006.1; JOINED.
DR EMBL; AF219929; AAF43006.1; JOINED.
DR EMBL; AF219930; AAF43006.1; JOINED.
DR EMBL; AF219931; AAF43006.1; JOINED.
DR EMBL; AF219932; AAF43006.1; JOINED.
DR EMBL; AF219933; AAF43006.1; JOINED.
DR EMBL; AF219934; AAF43006.1; JOINED.
DR EMBL; AF219935; AAF43006.1; JOINED.
DR EMBL; AF219936; AAF43006.1; JOINED.
DR EMBL; AF219937; AAF43006.1; JOINED.
DR EMBL; AF219938; AAF43006.1; JOINED.
DR Genew; HGNC:2855; DGK.
DR MIM; 604072; .
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000756; DAGKA.
DR InterPro; IPR001206; DAGKC.
DR InterPro; IPR002219; DAG_PE-bind.
DR InterPro; IPR003622; DAG_kin_cat.
DR Pfam; PF00023; ank; 2.
DR Pfam; PF00609; DAGKA; 1.
DR Pfam; PF00781; DAGKC; 1.
DR ProDom; PD002939; DAGKA; 1.
DR ProDom; PD005043; DAG_kin_cat; 1.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00109; CI; 2.
DR SMART; SM00045; DAGKA; 1.
DR SMART; SM00046; DAGKC; 1.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS50297; ANK_REPEAT; 1.
DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; FALSE_NEG.
DR PROSITE; PS50081; DAG_PE_BIND_DOM_2; FALSE_NEG.
DR Transferrase; Kinase; ANK repeat; Repeat; Nuclear protein;
KW Multimeric family; Polymorphism.
FT DOMAIN 178 232 PHORBOL-ESTER AND DAG BINDING 1.
FT DOMAIN 251 309 PHORBOL-ESTER AND DAG BINDING 2.
FT DOMAIN 339 353 MARKS HOMOLOGY REGION.
FT DOMAIN 374 500 CATALYTIC-A (POTENTIAL).
FT DOMAIN 526 683 CATALYTIC-B (POTENTIAL).
FT REPEAT 958 990 ANK 1.
FT REPEAT 997 1026 ANK 2.
FT DOMAIN 20 31 POLY-ALA.
FT DOMAIN 69 74 POLY-SER.
FT DOMAIN 95 102 POLY-ALA.
FT VARIANT 153 153 L -> F.
FT CONFLICT 160 160 A -> P (IN REF. 2).
SQ SEQUENCE 1065 AA; 116996 MW; B8497IAA7630A799 CRC64;
Query Match 47.7%; Score 41; DB 1; Length 1065;
Best Local Similarity 58.3%; Pred. No. 29;
```

```
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 5 RSNTYEVKGRDV 16
DB 1050 RQNYKVIGHEDL 1061
RESULT 14
LON_BORBU STANDARD; PRT; 806 AA.
ID LON_BORBU
AC Q59185;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP-dependent protease La (EC 3.4.21.53).
GN LON OR BB0253.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=97409968; PubMed=9266683;
RA Cloud J.L., Marconi R.T., Eggers C.H., Garon C.F., Tilly K.,
RA Samuels D.S.;
RT "Cloning and expression of the Borrelia burgdorferi lon gene.";
RL Gene 194:137-141(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
burgdorferi.";
RL Nature 390:580-586(1997).
CC -1- FUNCTION: DEGRADES SHORT-LIVED REGULATORY AND ABNORMAL PROTEINS IN
CC PRESENCE OF ATP. HYDROLYSES TWO ATPs FOR EACH PEPTIDE BOND CLEAVED
CC IN THE PROTEIN SUBSTRATE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of large proteins such as globin,
CC casein and denaturated serum albumin, in presence of ATP.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S16.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L77216; AAB72011.1; .
CC EMBL; AE001135; AAB91493.1; .
CC MEROPS; S16 UPW; .
CC TIGR; BB0253; .
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003959; AAA_ATPase_cent.
CC InterPro; IPR003111; LON.
CC InterPro; IPR001984; Lon_endopep.
CC InterPro; IPR004815; Lon_fam.
CC Pfam; PF00004; AAA; 1.
CC Pfam; PF02190; LON; 1.
CC PRINTS; PR00830; ENDOLAPTASE.
CC SMART; SM00382; AAA; 1.
CC SMART; SM00464; LON; 1.
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM.protein - protein search, using sw model

Run on: November 4, 2002, 19:54:31 ; Search time 94 Seconds  
(without alignments)  
35.072 Million cell updates/sec

Title: US-09-462-625-2\_COPY\_145\_160

Perfect score: 86

Sequence: 1 RGLRSNYEVKGHRDV 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	81	94.2	183	11 Q9JLN4	Q9JLN4 rattus norv
2	73	84.9	193	6 Q9GK12	Q9GK12 camelus dro
3	72	83.7	190	6 Q8SPP7	Q8SPP7 bos taurus
4	50	58.1	148	11 Q9DAL7	Q9DAL7 mus musculu
5	49	57.0	190	5 Q9VV97	Q9VV97 drosophila
6	47	54.7	978	5 Q905P6	Q905P6 haemochus
7	47	54.7	998	17 Q96252	Q96252 sulfolobus
8	46	53.5	215	17 Q82XU1	Q82XU1 pyrobaculum
9	46	53.5	432	2 Q93RK7	Q93RK7 bacillus sp
10	45	52.3	368	4 Q9HD75	Q9HD75 homo sapien
11	45	52.3	373	4 Q96LB8	Q96LB8 homo sapien
12	45	52.3	471	2 Q48204	Q48204 haemophilus
13	45	52.3	1478	5 Q9VB18	Q9VB18 drosophila
14	44.5	50.4	17	O57984	O57984 pyrococcus
15	44	51.2	341	4 Q96LB9	Q96LB9 homo sapien
16	44	51.2	460	2 Q48159	Q48159 haemophilus

17	44	51.2	1764	2	Q93T34	Q93T34 haemophilus
18	43	50.0	217	15	Q85497	Q85497 caprine art
19	43	50.0	365	10	Q9M471	Q9M471 dendrobium
20	43	50.0	370	10	Q9LW48	Q9LW48 drosanthemu
21	43	50.0	965	10	Q23946	Q23946 gossypium h
22	42	48.8	68	17	Q8TP58	Q8TP58 methanosarc
23	42	48.8	339	10	Q8S569	Q8S569 vitis vinif
24	42	48.8	500	11	Q9QX21	Q9QX21 mus musculu
25	42	48.8	529	11	Q9QX22	Q9QX22 mus musculu
26	42	48.8	530	11	Q8VCS0	Q8VCS0 mus musculu
27	42	48.8	674	10	Q8S915	Q8S915 nicotiana s
28	42	48.8	750	10	Q8S917	Q8S917 nicotiana s
29	42	48.8	787	16	Q9X1W8	Q9X1W8 thermotoga
30	42	48.8	960	10	Q9SA26	Q9SA26 zea mays (m
31	42	48.8	964	10	Q9SCB3	Q9SCB3 lycopersico
32	41	47.7	185	5	Q95S09	Q95S09 lycopersico
33	41	47.7	185	5	Q9V3B7	Q9V3B7 drosophila
34	41	47.7	712	3	Q8X0K3	Q8X0K3 neurospora
35	41	47.7	736	2	Q53042	Q53042 helicobacte
36	41	47.7	757	10	Q49985	Q49985 arabidopsis
37	41	47.7	757	10	P92977	P92977 arabidopsis
38	41	47.7	757	10	Q9C5W3	Q9C5W3 arabidopsis
39	41	47.7	830	10	Q64551	Q64551 arabidopsis
40	41	47.7	971	5	Q967C6	Q967C6 haemochus
41	40.5	47.1	318	16	Q9RRV3	Q9RRV3 deinococcus
42	40.5	47.1	469	16	O51751	O51751 borrelia bu
43	40.5	47.1	652	5	Q93346	Q93346 caenorhabdi
44	40	46.5	265	10	Q94ID8	Q94ID8 oryza sativ
45	40	46.5	328	10	O48623	O48623 triticum ae

ALIGNMENTS

RESULT 1

Q9JLN4 PRELIMINARY; PRT; 183 AA.  
ID Q9JLN4  
AC Q9JLN4  
DT 01-OCT-2000 (TREMREL. 15, Created)  
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)  
DT 01-OCT-2000 (TREMREL. 15, Last annotation update)  
DE Peptidoglycan recognition protein PGRP.  
GN PGRP.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=SPLEEN;  
RA Rehman A., Teodecki E.E., Krueger J.M.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF154114; AAF73252.1;  
SQ SEQUENCE 183 AA; 20590 MW; 5B9C1B7AA8A2EC21 CRC64;

Query Match 94.2%; Score 81; DB 11; Length 183;  
Best Local Similarity 100.0%; Pred. No. 3.3e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GFLRSNYEVKGHRDV 16  
|||||  
DB 147 GFLRSNYEVKGHRDV 161

RESULT 2

Q9GK12 PRELIMINARY; PRT; 193 AA.  
ID Q9GK12  
AC Q9GK12  
DT 01-MAR-2001 (TREMREL. 16, Created)  
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)  
DT 01-MAR-2002 (TREMREL. 20, Last annotation update)  
DE Peptidoglycan recognition protein precursor.  
GN PGRP.

```

OS Camelus dromedarius (Dromedary) (Arabic camel).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX NCBI_TaxID=9838;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LACTATING MAMMARY GLAND;
RA Kappeler S.R., Farah Z., Puhan Z.;
RT "Wilk as a Source of Camel (Camelus dromedarius) Peptidoglycan
RL Recognition Protein.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Kappeler S.R., Farah Z., Puhan Z.;
RT "The peptidoglycan recognition protein is expressed in the lactating
RL mammary gland of camels and binds to lactic acid bacteria.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ131676; CAC19553.1; -
DR EMBL; AJ409286; CAC84130.1; -
KW Signal.
FT SIGNAL.
FT CHAIN
SQ SEQUENCE 193 AA; 21377 MW; B6A1BD818030A7CB CRC64;
      1 21 POTENTIAL.
      22 193 PEPTIDOGLYCAN RECOGNITION PROTEIN.
      193 AA; 21377 MW; B6A1BD818030A7CB CRC64;

Query Match      84.9%; Score 73; DB 6; Length 193;
Best Local Similarity 93.3%; Pred. No. 8.8e-05;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GFRLRSNYEVKGRDV 16
Db 157 GFLRSNYEVKGRDV 171

RESULT 3
Q8SP7 ID Q8SP7 PRELIMINARY; PRT; 190 AA.
AC Q8SP7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Oligosaccharide-binding protein.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Tydel C.C., Yount N.Y., Tran D., Yuan J., Selsted M.E.;
RT "Isolation, characterization, and antimicrobial properties of bovine
RT oligosaccharide-binding protein: A microbicidal granule protein of
RT eosinophils and neutrophils.";
RL J. Biol. Chem. 0:0-0(2002).
DR EMBL; AY083309; AAL87002.1; -
SQ SEQUENCE 190 AA; 21063 MW; 2BA7D659438F4ED7 CRC64;

Query Match      83.7%; Score 72; DB 6; Length 190;
Best Local Similarity 81.2%; Pred. No. 0.00013;
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGFLRSNYEVKGRDV 16
Db 152 RGYLTPNYEVKGRDV 167

RESULT 4
Q9DAL7 ID Q9DAL7 PRELIMINARY; PRT; 148 AA.
AC Q9DAL7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 1700007L12RIK protein.

```

```

GN 1700007L12RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Rono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Borls A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK005734; BAB24212.1; -
DR MGD; MGI:1922720; 1700007L12RIK.
DR InterPro; IPR000048; IQ_region.
DR Pfam; PF00612; IQ; 2.
DR SMART; SM00015; IQ; 2.
SQ SEQUENCE 148 AA; 17809 MW; DIA416817431FB9 CRC64;

Query Match      58.1%; Score 50; DB 11; Length 148;
Best Local Similarity 57.1%; Pred. No. 0.69;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGFLRSNYEVKGRH 14
Db 109 RGFTQGNVEIKENR 122

RESULT 5
Q9VV97 ID Q9VV97 PRELIMINARY; PRT; 190 AA.
AC Q9VV97;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE CG9681 protein.
GN CGRP-SBL OR CG9681.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

```

RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier A.S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Galbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpén G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*."  
RL Science 287:2185-2195(2000).  
DR EMBL; AF003526; AAF49420.1;  
DR HSSP; P00806; IARO.  
DR FLYBase; FBgn0043578; PGRP-SB1.  
SQ SEQUENCE 190 AA; 20997 MW; A2F81E4903E9173E CRC64;

Query Match 57.0%; Score 49; DB 5; Length 190;  
Best Local Similarity 57.1%; Pred. No. 1.4;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RGFLRSNYEVKGHR 14

Db 152 RGYLKNYTLFGHR 165

RESULT 6

Q9U5P6 PRELIMINARY; PRT; 978 AA.  
AC Q9U5P6;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Microsomal aminopeptidase.  
GN H11-1.  
OS Haemochus contortus (Barber pole worm).  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;  
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.  
OX NCBI\_TaxID=6289;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Graham M., Smith T.S., Munn E.A., Coadwell W.J., Newton S.E.,  
RA Barker P.J.;  
RT "Cloning and sequencing of variants of H11, the highly protective  
RT membrane glycoprotein from *Haemonchus contortus*."  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ249941; CAB57357.1;  
DR MEROPS; M01.015;  
DR InterPro; IPR001930; Ala\_peptase.  
DR InterPro; IPR000130; Zn\_MTpeptidase.  
DR Pfam; PF01433; Peptidase.M1.1.  
DR PRINTS; PR00756; ALADIP7ASE.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
KW Aminopeptidase.  
SQ SEQUENCE 978 AA; 112117 MW; D4615283D35AB8E7 CRC64;

Query Match 54.7%; Score 47; DB 5; Length 978;

Best Local Similarity 50.0%; Pred. No. 18;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 RGFLRSNYEVKGHRDV 16  
Db 630 RAFCRSNYDANGWRNI 645  
RESULT 7  
Q96Z52 PRELIMINARY; PRT; 998 AA.  
AC Q96Z52;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein ST1979.  
GN ST1979.  
OS Sulfolobus tokodaii.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
OX NCBI\_TaxID=111955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-JCM 10545 / 7;  
RX PubMed-11572479;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,  
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
RA Yoshizawa T., Tanaka T., Kudooh Y., Yamazaki J., Kushida N., Oguchi A.,  
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
RA Oshima T., Kikuchi H.;  
RT "Complete genome sequence of an aerobic thermoacidophilic  
RT Crenarchaeon, *Sulfolobus tokodaii* strain 7."  
RL DNA Res. 8:123-140(2001).  
DR EMBL; AP000988; BAB67074.1;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 998 AA; 115169 MW; 350A01B763502366 CRC64;

Query Match 54.7%; Score 47; DB 17; Length 998;  
Best Local Similarity 62.5%; Pred. No. 19;  
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RGFLRSNYEVKGHRDV 16

Db 602 RGLLENPEKGVKDV 617

RESULT 8

Q8ZXU1 PRELIMINARY; PRT; 215 AA.  
AC Q8ZXU1;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein PAE1101.  
GN PAE1101.  
OS Pyrobaculum aerophilum.  
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
OC Thermoproteaceae; Pyrobaculum.  
OX NCBI\_TaxID=13773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-IM2 / ATCC 51768 / DSM 7523;  
RX PubMed-11792869;  
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
RA Miller J.H.;  
RT "Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum*  
RT aerophilum".  
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).  
DR EMBL; AE009803; AAL63255.1;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 215 AA; 24413 MW; 08CF10DA514F4363 CRC64;

Query Match 53.5%; Score 46; DB 17; Length 215;  
Best Local Similarity 60.0%; Pred. No. 5.2;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 GFRLSNYEKGHRDV 16  
||| |||||  
DB 87 GFLELRYEKGGRREV 101

RESULT 9  
Q93RK7

ID Q93RK7 PRELIMINARY; PRT; 432 AA.  
AC Q93RK7;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Replication protein.  
GN REP 43.  
OS Bacillus sp. KSM-KP43.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Bacillaceae; Bacillus.  
OX NCBI\_TaxID=109322;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KSM-KP43;  
RA Hatada Y., Ito S.;  
RT "A novel plasmid isolated from an alkaline Bacillus strain."  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB034994; BAB58973.1;  
SQ SEQUENCE 432 AA; 50516 MW; 27ADF6D62C1C1182 CRC64;

Query Match 53.5%; Score 46; DB 2; Length 432;  
Best Local Similarity 70.0%; Pred. No. 11;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 7 NYEVKGHRDV 16  
||:|||||:  
DB 247 NYQVKGHREL 256

RESULT 10  
Q9HD75

ID Q9HD75 PRELIMINARY; PRT; 368 AA.  
AC Q9HD75;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE Hypothetical 40.0 kDa protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Zhang W., Wan T., Cao X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF242518; AAF99599.1;  
DR InterPro; IPR002086; Aldehyde\_dehydr.  
DR PROSITE; PS00070; ALDEHYDE\_DEHYDR\_CYS; UNKNOWN\_1.  
KW Hypothetical protein.  
SQ SEQUENCE 368 AA; 40020 MW; 1E74970732A5DAFD CRC64;

Query Match 52.3%; Score 45; DB 4; Length 368;  
Best Local Similarity 50.0%; Pred. No. 14;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 RGFLRSNYEKGHRDV 16  
:|:| ||: |||  
DB 331 KGFLTPNLLVGHSDV 346

RESULT 11  
Q95LBB

ID Q96LBB PRELIMINARY; PRT; 373 AA.  
AC Q96LBB;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Peptidoglycan recognition protein-I-beta precursor.  
GN PGLYRPIBETA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21433985; PubMed=11461926;  
RA Liu C., Xu Z., Gupta D., Dziarski R.;  
RT "Peptidoglycan recognition proteins, a novel family of four human  
in innate immunity pattern recognition molecules.";  
J. Biol. Chem. 276.34686-34694(2001).  
RL EMBL; AY035377; AAK72485.1;  
DR InterPro; IPR002086; Aldehyde\_dehydr.  
DR PROSITE; PS00070; ALDEHYDE\_DEHYDR\_CYS; UNKNOWN\_1.  
KW Signal.  
FT SIGNAL  
SQ SEQUENCE 373 AA; 40662 MW; 1488A166018A66AA CRC64;

Query Match 52.3%; Score 45; DB 4; Length 373;  
Best Local Similarity 50.0%; Pred. No. 14;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 RGFLRSNYEKGHRDV 16  
:|:| ||: |||  
DB 336 KGFLTPNLLVGHSDV 351

RESULT 12  
Q48204

ID Q48204 PRELIMINARY; PRT; 471 AA.  
AC Q48204;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE IGA1 protease (Fragment).  
GN IGA.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HK284;  
RX MEDLINE=95302961; PubMed=7783620;  
RA Lomholt H., Poulsen K., Mogens K.;  
RT "Comparative characterization of the iga gene encoding IGA1 protease  
in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus  
influenzae.";  
Mol. Microbiol. 15:495-506(1995).  
RL EMBL; X82487; CAA57870.1;  
DR MEROPS; S06.001;  
DR InterPro; IPR00710; IGA\_S6.  
DR Pfam; PF02395; IGA1.  
KW Protease.  
FT NON\_TER  
FT NON\_TER  
SQ SEQUENCE 471 AA; 52559 MW; 68B57D31F21E6561 CRC64;

Query Match 52.3%; Score 45; DB 2; Length 471;  
Best Local Similarity 61.5%; Pred. No. 18;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 FLRSNYEKGHRD 15  
|:| |||||  
DB 357 FFKGNYEKGKTD 369

```

RESULT 13
Q9VB18
ID Q9VB18 PRELIMINARY; PRT; 1478 AA.
AC Q9VB18;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG5586 protein.
GN CG5586.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.T., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merklov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Pittman K.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Palazzolo M., Plattner K., Saunders R.D.C., Scheeler F., Shen H.,
RA Reinert K., Remington K., Simpson M., Skupski M.P., Smith T.,
RA Shue B.C., Siden-Kiamos I., Stapleton M., Strong R., Sun E.,
RA Spier E., Spradling A.C., Turner R., Venter E., Wang A.H., Wang X.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT The genome sequence of Drosophila melanogaster.;
RL Science 287:2185-2195(2000)
DR EMBL: AE003761; AAF56725.1;
DR FlyBase: FBgn0039530; CG5586.
DR InterPro: IPR001496; SOCS.
DR InterPro: IPR000007; Tubby.
DR InterPro: IPR001680; WD40.
DR Pfam: PF01167; Tub; 1.
DR Pfam: PF04000; WD40; 2.
DR SMART: SM00253; SOCS; 1.
DR SMART: SM00320; WD40; 1.
DR PROSITE: PSS0082; WD_REPEATS_2; 1.
DR PROSITE: PSS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 1478 AA; 163019 MW; 9993ED7E2FDD0D0C6 CRC64;

Query Match 52.3%; Score 45; DB 5; Length 1478;
Best Local Similarity 63.6%; Pred. NO. 64;

Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 LRSNVEVKGR 14
|:|:|:|:|:|
Db 121 LRTNVLGRHR 131

RESULT 14
O57984 PRELIMINARY; PRT; 504 AA.
ID O57984;
AC O57984;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 504AA long hypothetical TLDD protein.
GN PH0246.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayasi Y., Savada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka K., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo H., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.;
RL DNA Res. 5:55-76(1998)
DR EMBL: AP000001; BAA29318.1; -
DR InterPro: IPR002510; PmbA_TlD.
DR Pfam: PF01523; PmbA_TlD; 1.
KW Complete proteome.
SQ SEQUENCE 504 AA; 55950 MW; 593F665FC61CE462 CRC64;

Query Match 51.7%; Score 44.5; DB 17; Length 504;
Best Local Similarity 62.5%; Pred. NO. 24;
Matches 10; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 2 GFLRSNVEKGR-RDV 16
|:|:|:|:|:|
Db 436 GYLVEGEIKGLRDV 451

RESULT 15
O96LB9 PRELIMINARY; PRT; 341 AA.
ID O96LB9;
AC O96LB9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Peptidoglycan recognition protein-I-alpha precursor.
GN PGLYRPIALPHA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21433985; PubMed=11461926;
RA Liu C., Xu Z., Gupta D., Dziarski R.;
RT Peptidoglycan recognition proteins, a novel family of four human
innate immunity pattern recognition molecules.;
RL J. Biol. Chem. 276:34686-34694(2001).
DR EMBL: AY035376; AAK72484.1; -
KW Signal.
FT SIGNAL.
SQ SEQUENCE 341 AA; 37611 MW; 8ADD5AA97B632076 CRC64;

```

Query Match 51.2%; Score 44; DB 4; Length 341;  
Best Local Similarity 53.3%; Pred. No. 19;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
QY 2 GFLRSNYEVKGRDV 16  
|:| | | | | | | |  
Db 305 GYLTPNYLLMGHSDV 319

Search completed: November 4, 2002, 20:36:52  
Job time : 97 secs



[illegible]



```

repeat_region /rpt_family="AluSx"
complement(2576..2746)
repeat_region /rpt_family="AluSg/x"
complement(2747..3044)
repeat_region /rpt_family="AluSx"
3048..3750
repeat_region /rpt_family="MER64"
3841..4060
repeat_region /rpt_family="MER64"
4294..4451
misc_feature /note="DDS similarity to overlapping ESTs:
N39722 yx92e04.r1 Homo sapiens cDNA clone 269214 5';
(256..438); 81% identity.-AA283147 ztl4d07.r1
NCI_CGAP_GCB1 Homo sapiens cDNA
ztl49d07.r1 Soares pregnant uterus NBHPU Homo sapiens cDNA
clone 505261 5'; (127..281); 97% identity.-N28707
yx66dl1.r1 Homo sapiens cDNA clone 266709 5'; (67..222);
98% identity.-AA142881 ztl49d07.s1 Soares pregnant uterus
NBHPU Homo sapiens cDNA clone 505261 3'; (437..275); 97%
identity.
complement(4789..5083)
repeat_region /rpt_family="AluSx"
5875..5972
misc_feature /note="DDS similarity to overlapping ESTs:
N39722 yx92e04.r1 Homo sapiens cDNA clone 269214 5';
(439..506); 71% identity.-AA143160 ztl49d07.r1 Soares
pregnant uterus NBHPU Homo sapiens cDNA clone 505261 5';
(282..379); 100% identity.-AA142881 ztl49d07.s1 Soares
pregnant uterus NBHPU Homo sapiens cDNA clone 505261 3';
(274..175); 95% identity.
5875..5898
misc_feature /note="DDS similarity to N28707 yx66dl1.r1 Homo sapiens
cDNA clone 266709 5'; (223..246); 100% identity."
5906..6036
misc_feature /note="predicted exon, program: graill2exons_human_1.3,
frame: 1, quality: excellent, score: 100.000"
complement(6148..6443)
repeat_region /rpt_family="AluSx"
repeat_region complement(6608..6648)
repeat_region /rpt_family="LINE2"
complement(7118..7340)
repeat_region /rpt_family="MIR"
complement(7538..7639)
repeat_region /rpt_family="(CA)n"
complement(7792..8054)
repeat_region /rpt_family="MIR"
complement(8789..9083)
repeat_region /rpt_family="AluJo"
9456..9520
repeat_region /rpt_family="MER5B"
complement(9539..9841)
repeat_region /rpt_family="AluSx"
9909..9944
repeat_region /rpt_family="MER5B"
10099..10401
repeat_region /rpt_family="AluSx"
complement(10587..10752)
repeat_region /rpt_family="AluJ"
complement(10753..11057)
repeat_region /rpt_family="AluSx"
complement(11065..11191)
repeat_region /rpt_family="AluJ"
11322..11619
repeat_region /rpt_family="AluSg"
11775..11814
repeat_region /rpt_family="AT-rich"
complement(11864..11885)
repeat_region /rpt_family="AT-rich"
complement(11914..11976)
repeat_region /rpt_family="AluS"
complement(12055..12088)
repeat_region /rpt_family="AT-rich"

repeat_region complement(12186..12327)
/rpt_family="L1M1"
repeat_region complement(12334..12384)
/rpt_family="LINE2"
misc_feature 12784..12994
/note="predicted exon, program: graill2exons_human_1.3,
frame: 1, quality: excellent, score: 92.000"
repeat_region 12847..12908
/rpt_family="GC-rich"
12895..12994
misc_feature /note="DDS similarity to AA765166 nz79c05.s1 NCI_CGAP_GCB1
Homo sapiens cDNA clone IMAGE:1301672; (222..123); 97%
identity."
13002..13032
repeat_region /rpt_family="GC-rich"
13080..13162
misc_feature /note="predicted exon, program: graill2exons_human_1.3,
frame: 2, quality: excellent, score: 100.000-DDS
similarity to AA765166 nz79c05.s1 NCI_CGAP_GCB1 Homo
sapiens cDNA clone IMAGE:1301672; (122..48); 100%
identity."
repeat_region complement(13799..13954)
/rpt_family="GC-rich"

Alignment Scores:
Pred. No.: 3.64 Length: 166500
Score: 64.00 Matches: 12
Percent Similarity: 87.50% Conservative: 2
Best Local Similarity: 75.00% Mismatches: 2
Query Match: 74.42% Indels: 0
DB: 9 Gaps: 0

US-09-462-625-2_COPY_145_160 (1-16) x AC007785 (1-166500)
QY 1 ArgGlyPheLeuArgSerAsnTyrGluVallyscGlyHisArgAspVal 16
Db 17092 CAGGAGCCCTGAGGTCACCACTATGCTCTCAAGAGCACCGGGATGTG 17045

RESULT 14
AC002319
LOCUS AC002319 46305 bp DNA linear PRI 03-FEB-1998
DEFINITION Homo sapiens chromosome 9q34, clone 70C11, complete sequence.
AC002319
ACCESSION AC002319.1 GI:2828782
VERSION AC002319.1
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 46305)
AUTHORS Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P. and
Lander,E.
TITLE Homo sapiens chromosome 9q34, clone 70C11
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 46305)
Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
Barna,N., Chang,A., Cooke,P., Daly,M.J., Devon,K., Dewar,K.,
Forrest,C., Gage,D., Geraigery,K., Guitau,G., Hagos,B., Huang,J.,
Jacotot,L., Lane,M., Lee,K., Mackenzie,J., Marquis,N.,
McDermott,J., Molla,M., Moloney,N., Morrow,J., Nachman,A.,
Naylor,J., O'Connor,T., Olotu,A., Peterson,K., Rollins,G.,
Spencer,J., Stilwell,J., Stone,C., Strickland,C., Sydney,K.,
Traish,A., Wilmer,F., Zemtseva,I. and Zody,M.
Direct Submission
TITLE Submitted (15-JUL-1997) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS 3 (bases 1 to 46305)
Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P.,
Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R.,
Boutwell,C., Byrne,S., Cantu,C., Castle,A., Cooke,P., Daly,M.J.,
Depayre,E., Devon,K., Dewar,K., Duketete,B., Etemadi,S.,
Ferreira,P., Forrest,C., Gage,D., Gardyna,S., Gensheimer,S.,

```

Geraigery, K., Gilmarin, T., Gray, D., Hagos, B., Harris, K., Horton, L., Howland, J.C., Hui, L., Jacotot, L., Linton, L., Mackenzie, J., Marquis, N., McEwan, P., McGurk, A., Meldrim, J., Molla, M., Morris, W., Morrow, J., Nachman, A., Naylor, J., O'Connor, T., Pavlin, B., Peterson, K., Ranganath, S., Riley, R., Roberts, D., Rollins, G., Rossello, R., Roy, A., Shyam, R., Soohoo, S., Stilwell, J., Stone, C., Strickland, C., Sydney, K., Tang, L., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, Y., Ye, W.J., Zemtseva, I., Zhao, J. and Zody, M.

**TITLE** Direct Submission  
**JOURNAL** Submitted (31-JAN-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**REFERENCE** 4 (bases 1 to 46305)  
**AUTHORS** Birren, B., Fasnath, K., McKernan, K., Nussbaum, C., Richardson, P., Lander, E., Allen, N., Baker, J., Balgwin, J., Barna, N., Beckerly, R., Boutwell, C., Byrne, S., Cantu, C., Castle, A., Cooke, P., Daly, M.J., DePuyre, E., Devon, K., Dewar, K., DuRette, B., Etemadi, S., Ferreira, P., Forrest, C., Gage, D., Gardyna, S., Gensheimer, S., Geraigery, K., Gilmarin, T., Gray, D., Hagos, B., Harris, K., Horton, L., Howland, J.C., Hui, L., Jacotot, L., Linton, L., Mackenzie, J., Marquis, N., McEwan, P., McGurk, A., Meldrim, J., Molla, M., Morris, W., Morrow, J., Nachman, A., Naylor, J., O'Connor, T., Pavlin, B., Peterson, K., Ranganath, S., Riley, R., Roberts, D., Rollins, G., Rossello, R., Roy, A., Shyam, R., Soohoo, S., Stilwell, J., Stone, C., Strickland, C., Sydney, K., Tang, L., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, Y., Ye, W.J., Zemtseva, I., Zhao, J. and Zody, M.

**TITLE** Direct Submission  
**JOURNAL** Submitted (03-FEB-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Feb 3, 1998 this sequence version replaced gi:2828227. The Staden databases, finishing information, and all chromatographic files used in the assembly of this clone are available from our anonymous ftp site.

All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

#### FEATURES

Location/Qualifiers

1. 46305  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="9"  
 /map="9q34"  
 /clone="70C11"  
 /clone\_lib="Dr. David Kwiatkowski"  
 .complement(2..375)  
 /rpt\_family="L1M4"  
 .complement(377..687)  
 /rpt\_family="AluSg"  
 .complement(689..937)  
 /rpt\_family="L1M4"  
 .complement(914..1046)  
 /rpt\_family="L1"  
 1047..1346  
 /rpt\_family="AluSg"  
 .complement(1352..1708)  
 /rpt\_family="L1"  
 .complement(1694..2198)  
 /rpt\_family="L1"  
 .complement(2217..2503)  
 /rpt\_family="AluJb"  
 3497..3797  
 /rpt\_family="AluSx"  
 .complement(4593..4891)  
 /rpt\_family="AluSx"  
 5674..5954  
 /rpt\_family="HERV9"  
 5982..6273  
 /rpt\_family="AluSg"  
 .complement(6042..6226)  
 /note="Single-stranded, dye terminator coverage."  
 6636..7097

/rpt\_family="HERVE"  
 7441..7595  
 /rpt\_family="HERV9"  
 7693..8184  
 /rpt\_family="MER57\_internal"  
 8188..8427  
 /note="Single-stranded, dye terminator coverage."  
 8223..8518  
 /rpt\_family="AluY"  
 8520..8703  
 /rpt\_family="MER57\_internal"  
 9027..9325  
 /rpt\_family="AluSg"  
 10000..10294  
 /rpt\_family="AluSx"  
 10825..11061  
 /rpt\_family="MER57\_internal"  
 11115..11213  
 /rpt\_family="MER57\_internal"  
 11214..11641  
 /rpt\_family="MER57A"  
 .complement(11952..12258)  
 /rpt\_family="AluJb"  
 12281..12563  
 /rpt\_family="AluSx"  
 .complement(13005..13135)  
 /rpt\_family="MIR"  
 .complement(13279..13508)  
 /note="Single-stranded coverage."  
 14259..14563  
 /rpt\_family="AluSg"  
 .complement(14633..14919)  
 /rpt\_family="AluJb"  
 .complement(14935..15243)  
 /rpt\_family="AluSg"  
 .complement(15554..15925)  
 /rpt\_family="MER39"  
 .complement(16077..16452)  
 /rpt\_family="MER21A"  
 .complement(16458..16569)  
 /rpt\_family="AluSg/x"  
 .complement(16570..16768)  
 /rpt\_family="MER21A"  
 16769..16802  
 /rpt\_family="(GA)n"  
 .complement(16802..16825)  
 /rpt\_family="(CA)n"  
 .complement(18397..18432)  
 /rpt\_family="MIR"  
 18723..18800  
 /rpt\_family="LINE2"  
 .complement(18796..18884)  
 /rpt\_family="MIR"  
 .complement(18891..18922)  
 /rpt\_family="(TGAA)n"  
 19100..19386  
 /rpt\_family="AluSx"  
 19387..19451  
 /rpt\_family="Alu"  
 .complement(19522..19627)  
 /rpt\_family="FLAM\_C"  
 19666..19706  
 /rpt\_family="(GA)n"  
 21177..21179  
 /note="There is some evidence for 4 G's here instead of 3."  
 21784..21810  
 /rpt\_family="GC-rich"  
 22344..22631  
 /rpt\_family="AluSg"  
 23221..23281  
 /rpt\_family="LINE2"  
 23414..23450

```

repeat_region      /rpt_family="(TGAA)P"
                    complement(24150..24450)
repeat_region      /rpt_family="AluX"
                    complement(24854..24883)
repeat_region      /rpt_family="AT-rich"
                    complement(24884..25173)
repeat_region      /rpt_family="AluX"
                    complement(25174..25199)
repeat_region      /rpt_family="AT-rich"
                    complement(25385..25664)
repeat_region      /rpt_family="AluJo"
                    complement(25921..26223)
repeat_region      /rpt_family="AluX"
                    complement(26236..26522)
repeat_region      /rpt_family="LMB3"
                    complement(26598..26903)
repeat_region      /rpt_family="AluJb"
                    complement(27024..27316)
repeat_region      /rpt_family="AluSc"
                    complement(27364..27497)
repeat_region      /rpt_family="MIR"

```

```

Alignment Scores:
Pred. No.:      114      Length:      46305
Score:          53.00    Matches:      7
Percent Similarity: 100.00%    Conservative: 7
Best Local Similarity: 50.00%    Mismatches: 0
Query Match:      61.63%    Indels:      0
DB:              9      Gaps:      0

```

```

US-09-462-625-2_COPY_145_160 (1-16) x AC002319 (1-46305)

```

```

QY 3 PheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16

```

```

Db 339 TATTAGAACCATTTCCAAATCAAAGGCATAGAGATTG 380

```

# RESULT 15

AL732364/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

-----

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; SW., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9> RP11-430N14 is from the library RP11-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACe3.6.

## FEATURES

source

1..87903

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="9"

/clone\_lib="RP11-430N14"

/clone\_lib="RP11-11.2"

BASE COUNT 23441 a 21236 c 22502 g 20724 t

ORIGIN

Alignment Scores:

Pred. No.: 230 Length: 87903

Score: 53.00 Matches: 7

Percent Similarity: 100.00% Conservative: 7

Best Local Similarity: 50.00% Mismatches: 0

Query Match: 61.63% Indels: 0

DB: 9 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x AL732364 (1-87903)

QY 3 PheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16

Db 12761 TATTAGAACCATTTCCAAATCAAAGGCATAGAGATTG 12720

Search completed: November 4, 2002, 19:54:28

Job time: 2804 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 4, 2002, 17:49:26 ; Search time 301 Seconds  
(without alignments)  
119.708 Million cell updates/sec

Title: US-09-462-625-2\_COPY\_145\_160  
Perfect score: 86  
Sequence: 1 RGLRSNYEVKGRDV 16

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL-frame\_p2n.model -DBV-xlp  
-Q/cgn2\_1/USPTO.spool/US09462625/runat\_04112002\_110247\_23117/app\_query.fasta\_1.199  
-DB-N\_Geneseq\_101002 -OFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462625 -OCN=1\_1\_79 -runat\_04112002\_110247\_23117 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_101002.\*

1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	86	100.0	549	20	AAZ21819	Mouse tag7 clone c
2	86	100.0	677	18	AAT78510	Murine granulocyte
3	72	83.7	688	18	AAT78509	Bovine granulocyte
4	64	74.4	697	21	AAC58104	Human PRO1269 nucl
5	64	74.4	697	21	AAA37082	Human PRO1269 (UNQ
6	64	74.4	697	22	AAF54356	DNA encoding prote
7	64	74.4	718	20	AAZ21820	Human tag7 clone c
8	64	74.4	726	21	AAA80613	Human Htag7 secret
9	64	74.4	749	21	AAA51719	Chondrosarcoma pep
10	63	73.3	285	21	AAA80662	Human secreted pro
11	55	64.0	60	24	ABN38508	Human spliced tran
12	54	62.8	279	24	ABN24626	Human ORF polynuc
13	49	57.0	573	23	ABL15529	Drosophila melanog
14	49	57.0	2609	23	ABL15556	Drosophila melanog
15	49	57.0	2641	23	ABL15528	Drosophila melanog
16	48	55.8	65	24	ABN54729	Mouse spliced tran
17	48	55.8	621	22	AAH67897	C glutamicum codin
18	48	55.8	744	22	AAF71222	Corynebacterium gl
19	48	55.8	4703	23	ABL02316	Drosophila melanog
20	48	55.8	349980	22	AAH68533	C glutamicum codin
21	47	716	716	22	AAH17131	Human ion channel-
22	47	54.7	1689	14	AAQ52498	Helminth aminopept
23	47	54.7	3006	14	AAQ52491	Helminth aminopept
24	47	54.7	3084	14	AAQ52500	Helminth aminopept
25	46	53.5	1513	23	AAH23948	DNA encoding novel
26	46	53.5	2263	22	AAH23948	Bacillus sp. KSM-K
27	46	53.5	2880	22	AAH70647	Human immune/haema
28	46	53.5	4480	18	AAH71260	Rat sarcoma virus
29	46	53.5	4480	21	AAZ60823	Nucleotide sequenc
30	45	52.9	18909	23	ABL07204	Drosophila melanog
31	45.5	52.9	1429	22	AAH29726	S cerevisiae apopt
32	45	52.3	392	22	AAH35677	Human colon cancer
33	45	52.3	726	24	ABA90336	Human polynucleoti
34	45	52.3	807	22	AAH66438	C glutamicum codin
35	45	52.3	1110	22	AAH04006	Human full length
36	45	52.3	1110	24	ABA90337	Human polynucleoti
37	45	52.3	1128	22	AAH04004	Human full length
38	45	52.3	1876	21	AAA51718	Wound healing tiss
39	45	52.3	3395	24	ABQ70967	Listeria monocytog
40	45	52.3	4437	23	ABL09679	Drosophila melanog
41	45	52.3	6913	23	ABL09678	Drosophila melanog
42	45	52.3	349980	22	AAH68528	C glutamicum codin
43	44	51.2	1092	23	AAH91788	DNA encoding novel
44	44	51.2	7135	22	AAH46423	Tumour suppressor
45	43	50.0	439	21	AAC39717	Zea mays DNA fragm

ALIGNMENTS

RESULT 1  
AAZ21819  
ID AAZ21819 standard; cDNA; 549 BP.

XX AAZ21819;  
XX AC  
XX AAZ21819;  
XX 18-MAY-1999 (first entry)  
XX Mouse tag7 clone coding sequence.

XX Tag7: tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma;  
XX melanoma; leukaemia; apoptosis inducer; mouse; ds.  
XX Mus sp.

XX WO9902686-A1.  
XX PN  
XX 21-JAN-1999.  
PD

```

XX PF 10-JUL-1998; 98WO-EP04287.
XX PD 11-JUL-1997; 97US-0893764.
XX PR (BOEH ) BOEHRINGER INGELHEIM INT GMBH.
XX PA Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;
XX PI WPI; 1999-120887/10.
XX DR P-PSDB; AAY00770.
XX DR
XX PT New nucleic acid encoding tag7 - used to inhibit tumour growth and
XX PT induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and
XX PT leukaemia
XX PS Claim 3; Fig 1; 138pp; English.
XX CC This sequence encodes the murine tag7 of the invention. Cells containing
XX CC the tag7 DNA sequence are used to express recombinant tag7. Tag7 is used
XX CC to produce and purify antibodies; to inhibit growth of mammalian tumours,
XX CC especially for treating carcinoma (of liver, ovary, breast, cervix, lung,
XX CC prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head
XX CC and neck, squamous cell carcinoma or teratocarcinoma), sarcoma (Kaposi's,
XX CC osteo- or fibro-sarcomas), melanoma or leukaemia; and as a molecular
XX CC weight marker. The tag7 polypeptide inhibits tumour growth and induces
XX CC apoptosis. The tag7 coding sequences are also useful as probes for gene
XX CC mapping and detection of tag7 gene expression, and as primers. Antibodies
XX CC against tag7 are used as reagents for detecting tag7; as an antagonist of
XX CC tag7; for isolating tag7 and therapeutically to inhibit or delay tumour
XX CC metastasis.
XX SQ Sequence 549 BP; 117 A; 161 C; 151 G; 120 T; 0 other;

Alignment Scores:
Pred. No.: 4,61e-07 Length: 549
Score: 86.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 20 Gaps: 0

US-09-462-625-2_COPY_145_160 (1-16) x AAX21819 (1-549)
Qy 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16
Db 433 CGGGGCTTCTGAGATCCACTATGAAGTCAAAAGGACACCGGGATGTG 480

RESULT 2
AAT78510
ID AAT78510 standard; cDNA; 677 BP.
XX AC AAT78510;
XX DT 18-FEB-1998 (first entry)
XX DE Murine granulocyte peptide A precursor cDNA.
XX KW Antimicrobial peptide; antibiotic; antibacterial; antifungal;
XX KW fungicide; antiprotozoa; protozoacide; antiviral; virucide;
XX KW murine granulocyte peptide A; MGP-A; preservative; sepsis;
XX KW endotoxaemia; mouse; ss.
XX OS Mus musculus.
XX FH Key Location/Qualifiers
XX FT CDS 36..581
XX FT /tag= a
XX FT /transl_except= (pos:561..563, aa:Trp)
XX FT mat_peptide 543..578
XX FT /tag= b
XX FT /product= MGP-A
XX

FN WO9729765-A1.
PD 21-AUG-1997.
PF 13-FEB-1997; 97WO-US02218.
PR 16-FEB-1996; 96US-0011834.
PA (REGC ) UNIV CALIFORNIA.
XX PI Selsted ME;
XX DR WPI; 1997-424753/39.
XX DR P-PSDB; AAW23723.
XX PT Antimicrobial (poly)peptide(s) from bovine and murine granulocytes -
XX PT useful therapeutically, as preservatives for food, in water
XX PT treatment and in agriculture
XX PS Claim 14; Fig 5; 56pp; English.
XX CC This cDNA sequence encodes the precursor (see AAW23723) of a novel,
XX CC claimed antimicrobial peptide from bovine neutrophils, designated
XX CC murine granulocyte peptide A or MGP-A (see AAW23723). It was
XX CC isolated from murine bone marrow cDNA using primers based on
XX CC bovine BGP-A cDNA (see AAT78509). The encoded MGP-A precursor
XX CC comprises signal peptide and propeptide sequences followed by 12
XX CC C-terminal amino acids corresponding to mature MGP-A. MGP-A and
XX CC BGP-A (see AAW23724), exhibit activity against Gram-positive and
XX CC Gram-negative bacteria, fungi and viruses, specifically
XX CC Staphylococcus aureus, Escherichia coli, Candida albicans,
XX CC Salmonella typhimurium and C. neoformans (claimed). They can be
XX CC used in human or veterinary medicine (particularly to treat
XX CC disorders associated with lipopolysaccharides, e.g. sepsis and
XX CC endotoxaemia) or as preservatives in food products or in water
XX CC supplies (claimed). They can also be applied to crops to reduce
XX CC post-harvest spoilage or expressed in transgenic plants to increase
XX CC their disease resistance. They have low immunogenicity.
XX SQ Sequence 677 BP; 146 A; 217 C; 171 G; 143 T; 0 other;

Alignment Scores:
Pred. No.: 5,93e-07 Length: 677
Score: 86.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 18 Gaps: 0

US-09-462-625-2_COPY_145_160 (1-16) x AAT78510 (1-677)
Qy 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16
Db 465 CGGGGCTTCTGAGATCCAACTATGAAGTCAAAAGGACACCGGGATGTG 512

RESULT 3
AAT78509
ID AAT78509 standard; cDNA; 688 BP.
XX AC AAT78509;
XX DT 18-FEB-1998 (first entry)
XX DE Bovine granulocyte peptide A precursor cDNA.
XX KW Antimicrobial peptide; antibiotic; antibacterial; antifungal;
XX KW fungicide; antiprotozoa; protozoacide; antiviral; virucide;
XX KW bovine granulocyte peptide A; BGP-A; preservative; sepsis;
XX KW endotoxaemia; cattle; ss.
XX OS Bos taurus.
XX FH Key Location/Qualifiers

```





XX Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;  
KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening;  
KW ss.  
XX  
XX OS  
XX Homo sapiens.  
PN WO200012708-A2.  
XX  
PD 09-MAR-2000.  
XX  
XX 01-SEP-1999; 99WO-US20111.  
PR 01-SEP-1998; 98US-0098716.  
PR 01-SEP-1998; 98US-0098749.  
PR 01-SEP-1998; 98US-0098750.  
PR 02-SEP-1998; 98US-0098803.  
PR 02-SEP-1998; 98US-0098821.  
PR 02-SEP-1998; 98US-0098843.  
PR 09-SEP-1998; 98US-0099536.  
PR 09-SEP-1998; 98US-0099596.  
PR 09-SEP-1998; 98US-0099598.  
PR 09-SEP-1998; 98US-0099602.  
PR 09-SEP-1998; 98US-0099642.  
PR 10-SEP-1998; 98US-0099741.  
PR 10-SEP-1998; 98US-0099754.  
PR 10-SEP-1998; 98US-0099763.  
PR 10-SEP-1998; 98US-0099792.  
PR 10-SEP-1998; 98US-0099808.  
PR 10-SEP-1998; 98US-0099812.  
PR 10-SEP-1998; 98US-0099815.  
PR 10-SEP-1998; 98US-0099816.  
PR 15-SEP-1998; 98US-0100385.  
PR 15-SEP-1998; 98US-0100388.  
PR 15-SEP-1998; 98US-0100390.  
PR 16-SEP-1998; 98US-0100584.  
PR 16-SEP-1998; 98US-0100627.  
PR 16-SEP-1998; 98US-0100661.  
PR 16-SEP-1998; 98US-0100662.  
PR 16-SEP-1998; 98US-0100664.  
PR 17-SEP-1998; 98US-0100683.  
PR 17-SEP-1998; 98US-0100684.  
PR 17-SEP-1998; 98US-0100710.  
PR 17-SEP-1998; 98US-0100711.  
PR 17-SEP-1998; 98US-0100919.  
PR 17-SEP-1998; 98US-0100930.  
PR 18-SEP-1998; 98US-0100848.  
PR 18-SEP-1998; 98US-0100849.  
PR 18-SEP-1998; 98US-0101014.  
PR 18-SEP-1998; 98US-0101068.  
PR 18-SEP-1998; 98US-0101071.  
PR 22-SEP-1998; 98US-0101279.  
PR 23-SEP-1998; 98US-0101471.  
PR 23-SEP-1998; 98US-0101472.  
PR 23-SEP-1998; 98US-0101474.  
PR 23-SEP-1998; 98US-0101475.  
PR 23-SEP-1998; 98US-0101476.  
PR 23-SEP-1998; 98US-0101477.  
PR 23-SEP-1998; 98US-0101479.  
PR 24-SEP-1998; 98US-0101738.  
PR 24-SEP-1998; 98US-0101741.  
PR 24-SEP-1998; 98US-0101743.  
PR 24-SEP-1998; 98US-0101915.  
PR 24-SEP-1998; 98US-0101916.  
PR 29-SEP-1998; 98US-0102207.  
PR 29-SEP-1998; 98US-0102240.  
PR 29-SEP-1998; 98US-0102307.  
PR 29-SEP-1998; 98US-0102330.  
PR 29-SEP-1998; 98US-0102331.  
PR 30-SEP-1998; 98US-0102484.  
PR 30-SEP-1998; 98US-0102487.  
PR 30-SEP-1998; 98US-0102570.  
PR 30-SEP-1998; 98US-0102571.

PR 01-OCT-1998; 98US-0102684.  
PR 01-OCT-1998; 98US-0102687.  
PR 02-OCT-1998; 98US-0102965.  
PR 06-OCT-1998; 98US-0103258.  
PR 06-OCT-1998; 98US-0103449.  
PR 07-OCT-1998; 98US-0103314.  
PR 07-OCT-1998; 98US-0103315.  
PR 07-OCT-1998; 98US-0103328.  
PR 07-OCT-1998; 98US-0103395.  
PR 07-OCT-1998; 98US-0103396.  
PR 07-OCT-1998; 98US-0103401.  
PR 08-OCT-1998; 98US-0103633.  
PR 08-OCT-1998; 98US-0103678.  
PR 08-OCT-1998; 98US-0103679.  
PR 08-OCT-1998; 98US-0103711.  
PR 14-OCT-1998; 98US-0104257.  
PR 20-OCT-1998; 98US-0104987.  
PR 20-OCT-1998; 98US-0105000.  
PR 20-OCT-1998; 98US-0105002.  
PR 21-OCT-1998; 98US-0105104.  
PR 22-OCT-1998; 98US-0105169.  
PR 22-OCT-1998; 98US-0105266.  
PR 26-OCT-1998; 98US-0105693.  
PR 26-OCT-1998; 98US-0105694.  
PR 27-OCT-1998; 98US-0105807.  
PR 27-OCT-1998; 98US-0105881.  
PR 27-OCT-1998; 98US-0105882.  
PR 27-OCT-1998; 98US-0106062.  
PR 28-OCT-1998; 98US-0106023.  
PR 28-OCT-1998; 98US-0106029.  
PR 28-OCT-1998; 98US-0106030.  
PR 28-OCT-1998; 98US-0106032.  
PR 28-OCT-1998; 98US-0106033.  
PR 29-OCT-1998; 98US-0106178.  
PR 29-OCT-1998; 98US-0106248.  
PR 29-OCT-1998; 98US-0106384.  
PR 30-OCT-1998; 98US-0108500.  
PR 30-OCT-1998; 98US-0106464.  
PR 30-NOV-1998; 98US-0106856.  
PR 03-NOV-1998; 98US-0106902.  
PR 03-NOV-1998; 98US-0106905.  
PR 03-NOV-1998; 98US-0106919.  
PR 03-NOV-1998; 98US-0106932.  
PR 03-NOV-1998; 98US-0106934.  
PR 10-NOV-1998; 98US-0107783.  
PR 17-NOV-1998; 98US-0108775.  
PR 17-NOV-1998; 98US-0108779.  
PR 17-NOV-1998; 98US-0108787.  
PR 17-NOV-1998; 98US-0108788.  
PR 17-NOV-1998; 98US-0108801.  
PR 17-NOV-1998; 98US-0108802.  
PR 17-NOV-1998; 98US-0108806.  
PR 17-NOV-1998; 98US-0108807.  
PR 17-NOV-1998; 98US-0108867.  
PR 17-NOV-1998; 98US-0108925.  
PR 18-NOV-1998; 98US-0108848.  
PR 18-NOV-1998; 98US-0108849.  
PR 18-NOV-1998; 98US-0108850.  
PR 18-NOV-1998; 98US-0108851.  
PR 18-NOV-1998; 98US-0108852.  
PR 18-NOV-1998; 98US-0108858.  
PR 18-NOV-1998; 98US-0108904.

(GETH ) GENENTECH INC.

Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;

WPI: 2000-237871/20.

P-PSDB: AAY99400.

New mammalian DNA sequences encoding transmembrane, receptor or  
secreted PRO polypeptides, useful for screening of potential peptide or  
small molecule inhibitors of the relevant receptor/ligand interactions

XX PS Claim 2; Fig 121; 773pp; English.

XX CC AAA37022 to AAA37144 encode the new isolated human transmembrane,

XX CC receptor or secreted PRO polypeptides given in AA99340 to AA99462. The

XX CC transmembrane and receptor PRO proteins can be used for screening of

XX CC potential peptide or small molecule inhibitors of the relevant

XX CC receptor/ligand interactions. The polypeptides and nucleotide sequences

XX CC encoding then have various industrial applications, including uses as

XX CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent

XX CC PCR primers and hybridisation probes used in the isolation of the PRO

XX CC polypeptides from the present invention.

XX SQ Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;

Alignment Scores:

Pred. No.: 0.00951 Length: 697

Score: 64.00 Matches: 12

Percent Similarity: 87.50% Conservative: 2

Best Local Similarity: 75.00% Mismatches: 2

Query Match: 74.42% Indels: 0

DB: 21 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x AAA37082 (1-697)

QY 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16

Db 497 CAGGAGCCCTGAGTCCAACTATGTGTCTCAAGGACACCGGGATGTG 544

RESULT 6

AAF54356

ID AAF54356 standard; DNA; 697 BP.

XX AC AAF54356;

XX DT 02-APR-2001 (first entry)

XX DE DNA encoding protein of the invention #61.

XX KW Secreted; transmembrane; gene therapy; ss.

XX OS Unidentified.

XX PN WO200078961-A1.

XX PD 28-DEC-2000.

XX PF 18-FEB-2000; 2000WO-US04342.

XX PR 23-JUN-1999; 99US-0141037.

XX PR 20-JUL-1999; 99US-0144758.

XX PR 26-JUL-1999; 99US-0145698.

XX PR 01-SEP-1999; 99WO-US20111.

XX PR 30-OCT-1999; 99US-0162506.

XX PR 30-NOV-1999; 99WO-US28313.

XX PR 02-DEC-1999; 99WO-US28551.

XX PR 16-DEC-1999; 99WO-US30095.

XX PR 05-JAN-2000; 2000WO-US00219.

XX PR 06-JAN-2000; 2000WO-US00376.

XX (GETH ) GENENTECH INC.

XX PA Baker KP, Botstein D, Desnovers L, Eaton DL, Ferrara N, Fong S;

XX PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;

XX PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;

XX PI Watanabe CK, Williams PM, Wood WI;

XX WPI; 2001-071395/08.

XX DR Secreted and transmembrane proteins and nucleic acids designated PRO,

XX PT useful as hybridization probes, in chromosome and gene mapping and gene

XX PT therapy -

PS Claim 2; Fig 121; 787pp; English.

XX CC The present invention relates to secreted and transmembrane proteins.

XX CC These proteins and the DNA encoding them may be used as hybridization

XX CC probes, in chromosome and gene mapping and in the generation of

XX CC anti-sense RNA and DNA. They may also be used to generate either

XX CC transgenic animals or knockout animals which are in turn useful for

XX CC development and screening of therapeutically useful reagents.

XX CC The nucleic acids may also be used in gene therapy.

XX SQ Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;

Alignment Scores:

Pred. No.: 0.00951 Length: 697

Score: 64.00 Matches: 12

Percent Similarity: 87.50% Conservative: 2

Best Local Similarity: 75.00% Mismatches: 2

Query Match: 74.42% Indels: 0

DB: 22 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x AAF54356 (1-697)

QY 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16

Db 497 CAGGAGCCCTGAGTCCAACTATGTGTCTCAAGGACACCGGGATGTG 544

RESULT 7

AAAX21820

ID AAX21820 standard; cDNA to mRNA; 718 BP.

XX AC AAX21820;

XX DT 18-MAY-1999 (first entry)

XX DE Human tag7 clone coding sequence.

XX KW Tag7; tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma;

XX KW melanoma; leukaemia; apoptosis inducer; human; ss.

XX OS Homo sapiens.

XX PN WO9902686-A1.

XX XX 21-JAN-1999.

XX PF 10-JUL-1998; 98WO-EP04287.

XX PR 11-JUL-1997; 97US-0893764.

XX PA (BOEH ) BOEHRINGER INGELHEIM INT GMBH.

XX PI Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;

XX WPI; 1999-120887/10.

XX P-PSDB; AAY00771.

XX New nucleic acid encoding tag7 - used to inhibit tumour growth and

XX induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and

XX leukaemia

XX Claim 11; Page 126-127; 138pp; English.

XX CC This sequence encodes the human tag7 of the invention. Cells containing

XX CC the tag7 DNA sequence are used to express recombinant tag7. Tag7 is used

XX CC to produce and purify antibodies; to inhibit growth of mammalian tumours,

XX CC especially for treating carcinoma (of liver, ovary, breast, cervix, lung,

XX CC prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head

XX CC and neck, squamous cell carcinoma or teratocarcinoma), sarcoma (Kaposi's,

XX CC osteo- or fibro-sarcomas), melanoma or leukaemia; and as a molecular

XX CC weight marker. The tag7 polypeptide inhibits tumour growth and induces

XX CC apoptosis. The tag7 coding sequences are also useful as probes for gene

XX CC mapping and detection of tag7 gene expression, and as primers. Antibodies

XX CC against tag7 are used as reagents for detecting tag7; as an antagonist of

CC tag7; for isolating tag7 and therapeutically to inhibit or delay tumour  
 CC metastasis.

XX  
 SQ Sequence 718 BP; 142 A; 251 C; 194 G; 131 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.00985 Length: 718  
 Score: 64.00 Matches: 12  
 Percent Similarity: 87.50% Conservative: 2  
 Best Local Similarity: 75.00% Mismatches: 2  
 Query Match: 74.42% Indels: 0  
 DB: 20 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x AAX21820 (1-718)

QY 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16  
 Db 524 CAGGGAGCCCTGAGGTCCAACTATGTCTCAAAGGACACCGGGATGTG 571

RESULT 8

AAA80613  
 ID AAA80613 standard; CDNA; 726 BP.

XX AC AAA80613;

XX 21-NOV-2000 (first entry)

XX Human Htag7 secreted protein gene #8.

XX Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;  
 KW antirheumatic, dermatological; antiproliferative; antiarteriosclerotic;  
 KW anticancer; vulnery; antiviral; antibacterial; antifungal;  
 KW immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;  
 KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;  
 KW Crohn's disease; nephritis; hyperproliferative disorder;  
 KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;  
 KW melanoma; lymphoma; wound healing; human; ss.

XX Homo sapiens.

XX WO200029435-A1.

XX 25-MAY-2000.

XX 27-OCT-1999; 99WO-US25031.

XX 28-OCT-1998; 98US-0105971.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;

XX Greene JM;

XX WPI; 2000-387742/33.

XX P-PSDB; AAB25583.

XX Isolated nucleic acid molecules encoding human secreted proteins are  
 PT used for the prevention, amelioration and treatment of autoimmune,  
 PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,  
 PT wounds, and infectious diseases -

XX Claim 1; Figure 34; 803pp; English.

XX The present invention relates to 12 secreted human proteins and the  
 CC nucleotide sequences encoding them. The polynucleotide sequences given  
 CC in AAA80606-A80623 encode the 12 secreted protein sequences given in  
 CC AAB2576-B25593. The human secreted proteins have various activities  
 CC dependent on the tissues in which they are expressed. Examples of the  
 CC activities of the proteins include: immunosuppressant;  
 CC anti-inflammatory; antiarthritic; antirheumatic, dermatological;  
 CC antiproliferative; antiarteriosclerotic; anticancer; vulnery;  
 CC antiviral; antibacterial; and antifungal activity. The proteins,  
 CC polypeptides, agonists and antagonists may be used to treat prevent

CC and/or diagnose various disease, disorders and conditions examples of  
 CC which include: immune disorders e.g. Addison's disease, rheumatoid  
 CC arthritis, dermatitis, and multiple sclerosis; inflammatory disorders  
 CC e.g. inflammatory bowel disease, Crohn's disease and nephritis;  
 CC hyperproliferative disorders such as paraproteinemia and purpura;  
 CC cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;  
 CC cancer e.g. melanoma and lymphoma. The proteins and polynucleotide  
 CC sequences may also be used in wound healing and the treatment of  
 CC infectious diseases. The human secreted protein gene #8 and protein  
 CC sequences are represented in sequences AAA80613 and AAB25583. Sequences  
 CC AAA80662-A80663 represent genes related to the secreted protein gene#8.  
 XX  
 SQ Sequence 726 BP; 161 A; 251 C; 188 G; 126 T; 0 other;

Alignment Scores:

Pred. No.: 0.00999 Length: 726  
 Score: 64.00 Matches: 12  
 Percent Similarity: 87.50% Conservative: 2  
 Best Local Similarity: 75.00% Mismatches: 2  
 Query Match: 74.42% Indels: 0  
 DB: 21 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x AAA80613 (1-726)

QY 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16

Db 503 CAGGGAGCCCTGAGGTCCAACTATGTCTCAAAGGACACCGGGATGTG 550

RESULT 9

AAA51719

ID AAA51719 standard; CDNA; 749 BP.

XX AC AAA51719;

XX 31-OCT-2000 (first entry)

XX Chondrosarcoma peptidoglycan recognition protein-like protein cDNA.

XX Peptidoglycan recognition protein-like protein; PGRP; PGRP-C; regulator;  
 KW chondrosarcoma; immunosuppressive; antibacterial; anti-apoptotic;  
 KW tumour necrosis factor; TNF; anti-infectious; mononuclear phagocyte;  
 KW inhibitor; protein co-ordinate data; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 55..645  
 FT /\*tag= a  
 FT /product= PGRP-like\_protein

FT sig\_peptide 55..117

FT /\*tag= b

FT mat\_peptide 118..642

FT /\*tag= c

XX WO200039327-A1.

XX 06-JUL-2000.

XX 22-DEC-1999; 99WO-US30736.

XX 23-DEC-1998; 98US-0113809.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Young PE, Olsen HS;

XX WPI; 2000-452414/39.

XX P-PSDB; AAY96964.

XX Polynucleotide encoding peptidoglycan recognition protein-like protein,  
 PT antibodies specific to it useful for preventing, treating conditions  
 PT e.g. endotoxin shock and auto-immune disorders and infections in mammal

PS Claim 1; Fig 3; 19lpp; English.

XX Novel human peptidoglycan recognition protein-like proteins (PGRP) expressed by keratinocytes (PGRP-K), wound healing tissues (PGRP-W) or chondrosarcoma tissue (PGRP-C) are disclosed. The PGRP DNA and proteins are useful for preventing, treating or ameliorating a medical condition in a mammal (claimed). PGRP is useful in augmenting the immune system in such areas as immune recognition, antigen presentation and immune system activation. Antibodies or antagonists directed against these proteins may be useful in reducing or eliminating disorders associated with tumor necrosis factor (TNF) and TNF-like cytokines, such as endotoxin shock and autoimmune disorders and for treating infectious diseases including silicosis, sarcoidosis and idiopathic pulmonary fibrosis.

XX Sequence 749 BP; 165 A; 259 C; 195 G; 130 T; 0 other;

Alignment Scores:

Pred. No.: 0.0104 Length: 749  
Score: 64.00 Matches: 12  
Percent Similarity: 87.50% Conservatives: 2  
Best Local Similarity: 75.00% Mismatches: 2  
Query Match: 74.42% Indels: 0  
DB: 21 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x AAA51719 (1-749)

QY 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16

Db 526 CAGGGAGCCCTGAGTCCAACTATGTCTCAAGGACACCGGATGTG 573

RESULT 10

AAA80662

ID AAA80662 standard; cDNA; 285 BP.

XX AAA80662;

DT 21-NOV-2000 (first entry)

DE Human secreted protein gene #8 related gene HBWBTB79R SEQ ID #115.

XX Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic; antirheumatic, dermatological; antiproliferative; antiarteriosclerotic; anticancer; vulnery; antiviral; antibacterial; antifungal; immune disorder; Addison's disease; rheumatoid arthritis; dermatitis; multiple sclerosis; inflammatory disorder; inflammatory bowel disease; Crohn's disease; nephritis; hyperproliferative disorder; cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer; melanoma; lymphoma; wound healing; human; ss.

OS Homo sapiens.

PN WO200029435-A1.

XX 25-MAY-2000.

XX 27-OCT-1999; 99WO-US25031.

XX 28-OCT-1998; 98US-0105971.

XX (HUMA-) HUMAN GENOME SCI INC.

XX NI J., Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;

PI Greene JM;

XX WPI; 2000-387742/33.

XX Isolated nucleic acid molecules encoding human secreted proteins are used for the prevention, amelioration and treatment of autoimmune, inflammatory, hyperproliferative and cardiovascular disorders, cancer, wounds, and infectious diseases -

XX Disclosure; Page 757; 803pp; English.

PS

XX

CC The present invention relates to 12 secreted human proteins and the nucleotide sequences encoding them. The polynucleotide sequences given in AAA80606-A80623 encode the 12 secreted protein sequences given in AAB25576-B25593. The human secreted proteins have various activities dependent on the tissues in which they are expressed. Examples of the activities of the proteins include: immunosuppressant; anti-inflammatory; antiarthritic; antirheumatic; dermatological; antiproliferative; antiarteriosclerotic; anticancer; vulnery; antiviral; antibacterial; and antifungal activity. The proteins, polypeptides, agonists and antagonists may be used to treat prevent and/or diagnose various disease, disorders and conditions examples of which include: immune disorders e.g. Addison's disease, rheumatoid arthritis, dermatitis, and multiple sclerosis; inflammatory disorders e.g. inflammatory bowel disease, Crohn's disease and nephritis; hyperproliferative disorders such as paraproteinemia and purpura; cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis; cancer e.g. melanoma and lymphoma. The proteins and polynucleotide sequences may also be used in wound healing and the treatment of infectious diseases. The human secreted protein gene #8 and protein sequences are represented in sequences AAA80613 and AAB25583. Sequences AAA80662-A80663 represent genes related to the secreted protein gene#8.

XX Sequence 285 BP; 71 A; 99 C; 62 G; 47 T; 6 other;

Alignment Scores:

Pred. No.: 0.00503 Length: 285  
Score: 63.00 Matches: 12  
Percent Similarity: 86.67% Conservatives: 1  
Best Local Similarity: 80.00% Mismatches: 2  
Query Match: 73.26% Indels: 0  
DB: 21 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x AAA80662 (1-285)

QY 2 GlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16

Db 52 GGGAGCCTGAGTCCAACTATGTCTCAAGGACACCGGATGTG 96

RESULT 11

ABN38508

ID ABN38508 standard; DNA; 60 BP.

XX ABN38508;

AC ABN38508;

XX 15-JUL-2002 (first entry)

DT Human spliced transcript detection oligonucleotide SEQ ID NO:11256.

DE Human; mouse; rat; splice transcript; detection; RNA transcript;

XX splice variant; transcriptome; oligonucleotide library; ss.

XX Homo sapiens.

XX WO200210449-A2.

XX 07-FEB-2002.

XX 20-JUL-2001; 2001WO-IB01903.

XX 28-JUL-2000; 2000US-221607P.

XX 02-MAY-2001; 2001US-287724P.

XX (COMP-) COMPUGEN INC.

XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;

XX WPI; 2002-257383/30.

XX New oligonucleotide libraries comprising oligonucleotides which selectively hybridize to mRNAs transcribed from a transcription unit of a genome, useful for detecting tissue-, pathology-, and developmental-specific genes -

PT

XX PS Example 1; SEQ ID 11256; 47pp; English.

XX CC The present invention describes oligonucleotide libraries for detecting

XX CC messenger RNAs that populate a (sub-)transcriptome, where the

XX CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple

XX CC transcription units that populate a genome. The library comprises

XX CC several oligonucleotides, each capable of hybridizing selectively to a

XX CC set of messenger RNAs transcribed from a given transcription unit of

XX CC the genome, which encodes one or more messenger RNA splice variants.

XX CC The oligonucleotide libraries are useful for detecting mRNAs from a

XX CC biological sample, in expression profiling studies, in qualitatively or

XX CC quantitatively characterizing the corresponding transcriptome, and in

XX CC detecting RNA transcripts and splice variants of human or animal

XX CC transcriptomes. The libraries may also be used as specialised mini

XX CC libraries to detect transcripts of a sub-transcriptome under a

XX CC particular biological or pathological state, and so allowing the

XX CC detection of tissue- and pathology-specific genes such as those genes

XX CC only expressed in specific tissue under a specific pathological

XX CC condition; to detect developmental specific genes; and to detect RNA

XX CC transcripts and splice variants of a transcriptome of a patient suffering

XX CC from a particular disorder. ABN27253 to ABN59589 represent

XX CC oligonucleotide sequences from rats, humans and mice, which are used in

XX CC the exemplification of the present invention.

XX CC N.B. The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 60 BP; 16 A; 18 C; 15 G; 11 T; 0 other;

Alignment Scores:

Pred. No.: 0.0258 Length: 60

Score: 55.00 Matches: 10

Percent Similarity: 91.67% Conservative: 1

Best Local Similarity: 83.93% Mismatches: 1

Query Match: 63.95% Indels: 0

DB: 24 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x ABN38508 (1-60)

QY 5 ArgSerAsnTyrGluValLysGlyHisArgAspVal 16

Db 1 AGGTCCTAATGTGCTCTCAAGGACACCGGGATGTG 36

RESULT 12

ABN24626/C

ID ABN24626 standard; cDNA; 279 BP.

XX AC ABN24626;

XX DT 24-JUN-2002 (first entry)

XX DE Human ORFX polynucleotide sequence SEQ ID NO:17729.

XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;

XX KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;

XX KW degenerative disorder; osteoarthritis; neurodegenerative disorder;

XX KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;

XX KW hypertension; hypothyroidism; cholesterol ester storage disease;

XX KW immune deficiency; immune disorder; infectious disease;

XX KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;

XX KW myasthenia gravis; gene; ss.

XX OS Homo sapiens.

XX PN WO200192523-A2.

XX PD 06-DEC-2001.

XX PF 29-MAY-2001; 2001WO-US10836.

XX PR 30-MAY-2000; 2000US-206132P.

XX PR 29-AUG-2000; 2000US-228716P.

XX (CURA-) CURAGEN CORP.

XX PI Shinkets RA, Leach MD;

XX DR WPI; 2002-106308/14.

XX PS P-PSDB; ABP08874.

XX PT Novel human polypeptides and polynucleotides useful for diagnosing,

XX PT preventing and treating cardiovascular disease, neurodegenerative,

XX PT hyperproliferative disorders and autoimmune disorders

XX PS Disclosure; SEQ ID 17729; 1037pp; English.

XX CC The present invention describes substantially purified human proteins

XX CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1

XX CC in the specification). ABN15762 to ABN27252 encode the human ORFX

XX CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for

XX CC treating or preventing a pathology associated with an ORFX-associated

XX CC disorder in humans, and in the manufacture of a medicament for treating a

XX CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide

XX CC sequences can be used in gene therapy. ORFX sequences can be used in the

XX CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,

XX CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,

XX CC osteoarthritis, neurodegenerative disorders, disorders related to organ

XX CC transplantation, cardiovascular diseases, diabetes mellitus, systemic

XX CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester

XX CC storage disease, various immune deficiencies and disorders, infectious

XX CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid

XX CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host

XX CC disease and autoimmune inflammatory eye disease. ORFX proteins are also

XX CC useful for treating burns, incisions, ulcers, for treating osteoporosis,

XX CC bone degenerative disorders, or periodontal disease, and for gut

XX CC protection or regeneration and treatment of lung or liver fibrosis,

XX CC reperfusion injury in various tissues and conditions resulting from

XX CC systemic cytokine damage.

XX CC N.B. The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 279 BP; 79 A; 73 C; 48 G; 79 T; 0 other;

Alignment Scores:

Pred. No.: 0.254 Length: 279

Score: 54.00 Matches: 9

Percent Similarity: 85.71% Conservative: 3

Best Local Similarity: 64.29% Mismatches: 2

Query Match: 62.79% Indels: 0

DB: 24 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x ABN24626 (1-279)

QY 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArg 14

Db 64 CGTGGCTTTATTCGGGGCACTATGAAATCAAAGAAACCGA 23

RESULT 13

ABN15529

ID ABL15529 standard; cDNA; 573 BP.

XX AC ABL15529;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 41069.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical; gene; ss.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

PD 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI; 2001-656860/75.  
DR P-PSDB; ABB71426.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Claim 1; SEQ ID NO 41069; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 573 BP; 136 A; 160 C; 147 G; 130 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 5.4 Length: 573  
Score: 49.00 Matches: 8  
Percent Similarity: 78.57% Conservative: 3  
Best Local Similarity: 57.14% Mismatches: 3  
Query Match: 56.98% Indels: 0  
DB: 23 Gaps: 0  
  
US-09-462-625-2\_COPY\_145\_160 (1-16) x ABL15529 (1-573)  
QY 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArg 14  
|||||:|||||:|||||:|||||:|||||  
DB 454 CGTGGATACCTCAAGGATACTACACGCTGTTCGGTCATCGG 495  
  
RESULT 14  
ABL15556  
ID ABL15556 standard; cDNA; 2609 BP.  
XX  
XX ABL15556;  
XX  
XX 26-MAR-2002 (first entry)  
DT  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 41150.  
DE  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; ss.  
KW  
XX Drosophila melanogaster.  
OS  
XX WO200171042-A2.  
PN  
XX 27-SEP-2001.  
PD  
XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
PA

XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI; 2001-656860/75.  
DR P-PSDB; ABB71453.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Claim 1; SEQ ID NO 41150; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2609 BP; 711 A; 606 C; 587 G; 705 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 33.4 Length: 2609  
Score: 49.00 Matches: 8  
Percent Similarity: 78.57% Conservative: 3  
Best Local Similarity: 57.14% Mismatches: 3  
Query Match: 56.98% Indels: 0  
DB: 23 Gaps: 0  
  
US-09-462-625-2\_COPY\_145\_160 (1-16) x ABL15556 (1-2609)  
QY 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArg 14  
|||||:|||||:|||||:|||||:|||||  
DB 695 CGTGGATACCTCAAGGATACTACACGCTGTTCGGTCATCGG 736  
  
RESULT 15  
ABL15528  
ID ABL15528 standard; cDNA; 2641 BP.  
XX  
XX ABL15528;  
XX  
XX 26-MAR-2002 (first entry)  
DT  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 41066.  
DE  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; ss.  
KW  
XX Drosophila melanogaster.  
OS  
XX WO200171042-A2.  
PN  
XX 27-SEP-2001.  
PD  
XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
PA  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI; 2001-656860/75.  
DR P-PSDB; ABB71425.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

PT interactions -  
XX  
PS Claim 1; SEQ ID NO 41066; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2641 BP; 785 A; 568 C; 598 G; 690 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 33.9 Length: 2641  
Score: 49.00 Matches: 8  
Percent Similarity: 78.57% Conservative: 3  
Best Local Similarity: 57.14% Mismatches: 3  
Query Match: 56.98% Indels: 0  
DB: 23 Gaps: 0  
  
US-09-462-625-2\_COPY\_145\_160 (1-16) x ABL15528 (1-2641)  
  
Qy 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArg 14  
|||||:|||||: ||||| :|||  
Db 1522 CGTGGATACCTCAAGGATAACTACACGCTGTTTCGGTCATCGG 1563

Search completed: November 4, 2002, 19:07:34  
Job time : 304 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 4, 2002, 19:00:46 ; Search time 83 Seconds  
(without alignments)

64.223 Million cell updates/sec

Title: US-09-462-625-2\_COPY\_145\_160

Perfect score: 86

Sequence: 1 RGFLRSNYEVKGRDV 16

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool/US09462625/runat\_04112002\_110249\_23150/app\_query.fasta\_1.199  
-DB=PublishedApplications\_NA -OFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40 cdi -LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09462625\_RCGN\_1\_1\_36\_erunat\_04112002\_110249\_23150  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications\_NA.\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	44	51.2	684973	10	US-09-263-959-1 Sequence 1, Appl1
2	43	50.0	550	10	US-09-864-761-14908 Sequence 14908, A
3	43	50.0	2228	10	US-09-822-830A-424 Sequence 424, App
C 4	43	50.0	4377	12	US-10-054-691-1 Sequence 1, Appl1

5	43	50.0	17216	10	US-09-764-877-3565	Sequence 3565, Ap
6	43	50.0	17217	10	US-09-764-877-3566	Sequence 3566, Ap
C 7	42	48.8	307	10	US-09-880-253A-46	Sequence 46, Appl
C 8	42	48.8	35100	10	US-09-782-378A-26	Sequence 26, Appl
9	41	47.7	143	10	US-09-983-965-2238	Sequence 2238, Ap
10	41	47.7	454	10	US-09-998-998-1440	Sequence 1440, Ap
11	41	47.7	472	10	US-09-770-444-152	Sequence 152, App
12	41	47.7	1137	10	US-09-995-494-21	Sequence 152, App
13	41	47.7	2081	10	US-09-771-161A-57	Sequence 57, Appl
14	41	47.7	2211	10	US-09-815-242-7196	Sequence 7196, Ap
15	41	47.7	2211	10	US-09-815-242-7376	Sequence 7376, Ap
C 16	41	47.7	170834	10	US-09-835-232-7	Sequence 7376, Ap
17	40	46.5	185	10	US-09-878-574-122	Sequence 122, Appl
18	40	46.5	256	10	US-09-878-574-5860	Sequence 5860, Ap
19	40	46.5	817	10	US-09-966-881-36	Sequence 36, Appl
20	40	46.5	1114	10	US-09-888-358-1	Sequence 1, Appl1
21	40	46.5	1546	10	US-09-888-358-2	Sequence 2, Appl1
22	40	46.5	1643	10	US-09-840-787-68	Sequence 68, Appl
C 23	40	46.5	1758	10	US-09-894-018-88	Sequence 88, Appl
C 24	40	46.5	2898	10	US-09-765-231A-41	Sequence 41, Appl
C 25	40	46.5	2933	12	US-10-044-090-367	Sequence 367, App
C 26	40	46.5	3008	10	US-09-925-301-350	Sequence 350, App
27	39.5	45.9	174424	10	US-09-967-768A-314	Sequence 314, App
C 28	39	45.3	273	10	US-09-960-352-1754	Sequence 1754, Ap
29	39	45.3	289	10	US-09-878-574-14787	Sequence 14787, A
C 30	39	45.3	473	10	US-09-917-800A-711	Sequence 711, App
C 31	39	45.3	480	10	US-09-864-761-6550	Sequence 6550, Ap
C 32	39	45.3	964	10	US-09-974-300-1894	Sequence 1894, Ap
C 33	39	45.3	1192	10	US-09-880-107-1592	Sequence 1592, Ap
C 34	39	45.3	2036	10	US-09-729-674-79	Sequence 79, Appl
C 35	39	45.3	3466	10	US-09-782-906-1	Sequence 1, Appl1
C 36	39	45.3	4839	12	US-10-016-358-1	Sequence 1, Appl1
37	39	45.3	143306	10	US-09-729-920-3	Sequence 3, Appl1
C 38	39	45.3	397658	10	US-09-813-320-3	Sequence 3, Appl1
39	38.5	44.8	181	10	US-09-563-817-461	Sequence 461, App
40	38	44.2	180	10	US-09-864-761-18235	Sequence 18235, A
41	38	44.2	207	10	US-09-974-300-5121	Sequence 5121, Ap
C 42	38	44.2	233	10	US-09-960-352-14489	Sequence 14489, A
43	38	44.2	234	10	US-09-867-701-8474	Sequence 8474, Ap
44	38	44.2	235	10	US-09-878-574-510	Sequence 510, App
45	38	44.2	385	10	US-09-878-574-1471	Sequence 1471, Ap

ALIGNMENTS

RESULT 1  
US-09-263-959-1/c  
; Sequence 1, Application US/09263959  
; Patent No. US20020150891A1  
; GENERAL INFORMATION:  
; APPLICANT: Hood, Leroy E.  
; APPLICANT: Koop, Ben F.  
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH  
; NUMBER OF SEQUENCES: 1279  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: US  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/263,959  
; FILING DATE: 05-MAR-1999  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mcmasters, David D.



```

; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 14908
; LENGTH: 550
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004832.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
US-09-864-761-14908

Alignment Scores:
Pred. No.: 11.7 Length: 550
Score: 43.00 Matches: 8
Percent Similarity: 76.92% Conservative: 2
Best Local Similarity: 61.54% Mismatches: 3
Query Match: 50.00% Indels: 0
DB: 10 Gaps: 0

US-09-462-625-2_COPY_145_160 (1-16) x US-09-864-761-14908 (1-550)

Qy 3 PheLeuArgSerAsnTyrGluValLysGlyHisArgAsp 15
: ::::::::::: ::::: ||:::::::::
Db 502 TACCTAAGGAGTCCCGAGGAGCTGGCTGGCCATCGTGAT 540

RESULT 3
US-09-822-830A-424
; Sequence 424, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
; APPLICANT: Genetics Institute, Inc.
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6402
; CURRENT APPLICATION NUMBER: US/09/822,830A
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,604
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 424
; LENGTH: 2228
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 2055,2076,2140,2171,2172,2174,2190
; OTHER INFORMATION: n=a,c,g, or t
US-09-822-830A-424

Alignment Scores:
Pred. No.: 63.5 Length: 2228
Score: 43.00 Matches: 7
Percent Similarity: 83.33% Conservative: 3
Best Local Similarity: 58.33% Mismatches: 2
Query Match: 50.00% Indels: 0
DB: 10 Gaps: 0

US-09-462-625-2_COPY_145_160 (1-16) x US-09-822-830A-424 (1-2228)

```

```
QY 5 ArgSerAsnTyrGluValLysGlyHisArgAspVal 16
||||: ||| |||:|||||:|||||:
Db 1640 AGGACAAATACACTGTGGAGGCCACAGGACCTC 1675

RESULT 4
US-10-054-691-1/c
; Sequence 1, Application US/10054691
; Patent No. US20020115846A1
; GENERAL INFORMATION:
; APPLICANT: Yu. Xuanchuan
; APPLICANT: Miranda, Maricar
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20020115846A1 Human Lipase and Polynucleotides Encoding t
; FILE REFERENCE: LEX-0303-USA
; CURRENT APPLICATION NUMBER: US/10/054, 691
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 60/264,049
; PRIOR FILING DATE: 2001-01-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4377
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-054-691-1

Alignment Scores:
Pred. No.: 144 Length: 4377
Score: 43.00 Matches: 7
Percent Similarity: 76.92% Conservative: 3
Best Local Similarity: 53.85% Mismatches: 3
Query Match: 50.00% Indels: 0
DB: 12 Gaps: 0

US-09-462-625-2_COPY_145_160 (1-16) x US-10-054-691-1 (1-4377)

QY 4 LeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16
||||: ||| |||:|||||:|||||:
Db 1769 TTATCATCAACATCTCAGGACACAGGGACAGACGACCTG 1731

RESULT 5
US-09-764-877-3565
; Sequence 3565, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3565
; LENGTH: 17216
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3565

Alignment Scores:
Pred. No.: 751 Length: 17216
Score: 43.00 Matches: 8
Percent Similarity: 76.92% Conservative: 2
Best Local Similarity: 61.54% Mismatches: 3
Query Match: 50.00% Indels: 0
DB: 10 Gaps: 0

US-09-462-625-2_COPY_145_160 (1-16) x US-09-764-877-3565 (1-17216)

QY 3 PheLeuArgSerAsnTyrGluValLysGlyHisArgAsp 15
||||: ||| |||:|||||:|||||:
Db 15183 TACCTAAGGAGTCCCGGAGGCTGGTGGCCATCGTGAT 15221
```

```
RESULT 6
US-09-764-877-3566
; Sequence 3566, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3566
; LENGTH: 17217
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3566

Alignment Scores:
Pred. No.: 752 Length: 17217
Score: 43.00 Matches: 8
Percent Similarity: 76.92% Conservative: 2
Best Local Similarity: 61.54% Mismatches: 3
Query Match: 50.00% Indels: 0
DB: 10 Gaps: 0

US-09-462-625-2_COPY_145_160 (1-16) x US-09-764-877-3566 (1-17217)

QY 3 PheLeuArgSerAsnTyrGluValLysGlyHisArgAsp 15
||||: ||| |||:|||||:|||||:
Db 15183 TACCTAAGGAGTCCCGGAGGCTGGTGGCCATCGTGAT 15221

RESULT 7
US-09-880-253A-46/c
; Sequence 46, Application US/09880253A
; Patent No. US2002011322A1
; GENERAL INFORMATION:
; APPLICANT: The University of Queensland
; TITLE OF INVENTION: EXPRESSION MODULATING SEQUENCES
; FILE REFERENCE: 2415281/EJH
; CURRENT APPLICATION NUMBER: US/09/880,253A
; CURRENT FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 46
; LENGTH: 307
; TYPE: RNA
; ORGANISM: mouse
US-09-880-253A-46

Alignment Scores:
Pred. No.: 8.94 Length: 307
Score: 42.00 Matches: 7
Percent Similarity: 75.00% Conservative: 5
Best Local Similarity: 43.75% Mismatches: 4
Query Match: 48.84% Indels: 0
DB: 10 Gaps: 0

US-09-462-625-2_COPY_145_160 (1-16) x US-09-880-253A-46 (1-307)

QY 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16
||||: ||| |||:|||||:|||||:
Db 93 AAGGGAGAGATTCTCTAGTATGTGGAGGCGCGGACATAGACATC 46

RESULT 8
US-09-782-378A-26/c
; Sequence 26, Application US/09782378A
; Patent No. US20020102731A1
; GENERAL INFORMATION:
; APPLICANT: Hearing, Patrick
; APPLICANT: Bahou, Wadie
; APPLICANT: Sandalon, Ziv
```

```
; APPLICANT: Gnatenko, Dmitri
; TITLE OF INVENTION: Adenoviral Vectors
; FILE REFERENCE: STONYB-04970
; CURRENT APPLICATION NUMBER: US/09/782,378A
; CURRENT FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/237,747
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 26
; LENGTH: 35100
; TYPE: DNA
; ORGANISM: Human adenovirus type 17
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (25184)..(25184)
; OTHER INFORMATION: The n at this position can be a, c, t, or g.
US-09-782-378A-26

Alignment Scores:
Pred. No.: 2,75e+03 Length: 35100
Score: 42.00 Matches: 8
Percent Similarity: 71.43% Conservative: 2
Best Local Similarity: 57.14% Mismatches: 4
Query Match: 48.84% Indels: 0
DB: 10 Gaps: 0

US-09-462-625-2_COPY_145_160 (1-16) x US-09-782-378A-26 (1-35100)
Qy 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArg 14
|||||
Db 8971 CGTGGATTCTACGAGCCATGGAAGCGTGGAGGCCATCGG 8930

RESULT 9
US-09-983-965-2238
; Sequence 2238, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 2238
; LENGTH: 143
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 50-LIB3057-002-Q1-K1-E6
US-09-983-965-2238

Alignment Scores:
Pred. No.: 5.49 Length: 143
Score: 41.00 Matches: 7
Percent Similarity: 90.00% Conservative: 2
Best Local Similarity: 70.00% Mismatches: 1
Query Match: 47.67% Indels: 0
DB: 10 Gaps: 0

US-09-462-625-2_COPY_145_160 (1-16) x US-09-983-965-2238 (1-143)
Qy 2 GlyPheLeuArgSerAsnTyrGluValLys 11
|||||
Db 73 GGCCTTATAGATCTACTACATCATATAA 102

RESULT 10
US-09-998-598-1440
; Sequence 1440, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Meagher, Madelein Joy
; APPLICANT: Chenault, Ruth A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 1440
; LENGTH: 454
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-1440

Alignment Scores:
Pred. No.: 22.2 Length: 454
Score: 41.00 Matches: 7
Percent Similarity: 71.43% Conservative: 3
Best Local Similarity: 50.00% Mismatches: 4
Query Match: 47.67% Indels: 0
DB: 10 Gaps: 0

US-09-462-625-2_COPY_145_160 (1-16) x US-09-998-598-1440 (1-454)
Qy 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArg 14
|||||
Db 102 CGTCCTTTCATCAATCACTGTGCTATTAATAAAGTCACCGT 143

RESULT 11
US-09-770-444-152
; Sequence 152, Application US/09770444
; Patent No. US20020023280A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kricker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2027 (PARA-016PRV)
; CURRENT APPLICATION NUMBER: US/09/770,444
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,502
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 152
; LENGTH: 472
; TYPE: DNA
```

Alignment Scores:	Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-770-444-152	23.2	41.00	472	8	3	0	0	0
US-09-462-625-2_COPY_145_160 (1-16) x US-09-770-444-152 (1-472)								
QY 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16								
Db 127 CGGGGTTCTTGGATAACCGTTTCGTGTTAAAGGTTGCAGTGATCTA 174								
US-09-995-494-21								
Sequence 21, Application US/09995494								
Patent No. US20020127578A1								
GENERAL INFORMATION:								
APPLICANT: Macina, Roberto								
APPLICANT: Recipon, Herve								
APPLICANT: Caferkey, Robert								
APPLICANT: Ali, Shujath								
APPLICANT: Sun, Yongming								
APPLICANT: Liu, Chenghua								
APPLICANT: Chen, Sei-yu								
TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes and								
FILE REFERENCE: DEX-0293								
CURRENT APPLICATION NUMBER: US/09/995,494								
CURRENT FILING DATE: 2001-11-27								
PRIOR APPLICATION NUMBER: 60/253,176								
PRIOR FILING DATE: 2000-11-27								
NUMBER OF SEQ ID NOS: 115								
SOFTWARE: PatentIn version 3.1								
SEQ ID NO 21								
LENGTH: 1137								
TYPE: DNA								
ORGANISM: Homo sapien								
US-09-995-494-21								
Alignment Scores:								
Pred. No.:	67.3		1137					
Score:	41.00		7					
Percent Similarity:	88.89%		Conservative:	1				
Best Local Similarity:	77.78%		Mismatches:	1				
Query Match:	47.67%		Indels:	0				
DB:	10		Gaps:	0				
US-09-462-625-2_COPY_145_160 (1-16) x US-09-995-494-21 (1-1137)								
QY 3 PheLeuArgSerAsnTyrGluValLys 11								
Db 224 TTCCTAAGGAAGAATTATGAATAAAA 250								
US-09-771-161A-57								
Sequence 57, Application US/09771161A								
Patent No. US20020110811A1								
GENERAL INFORMATION:								
APPLICANT: LEVINE, et al.								
TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES								
FILE REFERENCE: 802620-2005.1								
CURRENT APPLICATION NUMBER: US/09/771,161A								
CURRENT FILING DATE: 2001-01-26								
PRIOR APPLICATION NUMBER: 09/724,676								
PRIOR FILING DATE: 2000-11-28								
PRIOR APPLICATION NUMBER: 136776								
PRIOR FILING DATE: 2000-06-15								
PRIOR APPLICATION NUMBER: 135619								
US-09-462-625-2_COPY_145_160 (1-16) x US-09-771-161A-57 (1-2081)								
QY 5 ArgSerAsnTyrGluValLysGlyHisArgAspVal 16								
Db 443 CGTCAGAACTATAGCTCATTTGGCCATGAGGACCTG 478								
US-09-815-242-7196								
Sequence 7196, Application US/09815242								
Patent No. US20020061569A1				</				

```
US-09-462-625-2_COPY_145_160 (1-16) x US-09-815-242-7196 (1-2211)

QY 1 ArgGlyPheLeuArgSerAsnTyrGluVal-----LysGlyHis---ArgAspVal 16
   :: ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 49 AAAAATTTTGGATAAAATATGAAGTCATTCCTCTAAAGGCGATGTTAGGATTTA 108

RESULT 15
US-09-815-242-7376
; Sequence 7376, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7376
; LENGTH: 2211
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2211)
US-09-815-242-7376

Alignment Scores:
Pred. No.: 150 Length: 2211
Score: 41.00 Matches: 11
Percent Similarity: 65.00% Conservative: 2
Best Local Similarity: 55.00% Mismatches: 3
Query Match: 47.67% Indels: 4
DB: 10 Gaps: 2

US-09-462-625-2_COPY_145_160 (1-16) x US-09-815-242-7376 (1-2211)

QY 1 ArgGlyPheLeuArgSerAsnTyrGluVal-----LysGlyHis---ArgAspVal 16
   :: ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 49 AAAAATTTTGGATAAAATATGAAGTCATTCCTCTAAAGGCGATGTCAGGATTTA 108

Search completed: November 4, 2002, 20:33:21
Job time : 128 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 4, 2002, 17:50:51 ; Search time 2772 Seconds  
(without alignments)  
167.982 Million cell updates/sec

Title: US-09-462-625-2\_COPY\_145\_160  
Perfect score: 86  
Sequence: 1 RGLRSNYEVKGRDV 16

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlip  
-O/cgn2\_1/USPTO.spool/US09462625/runat\_04112002\_110248\_23125/app\_query.fasta\_1.199  
-DB=GenEmbl -OFFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462625 @CGN\_1\_1\_2659 @runa\_04112002\_110248\_23125 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sv.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*  
30: em.htg\_hum.\*  
31: em.htg\_inv.\*  
32: em.htg\_other.\*  
33: em.htg\_mus.\*  
34: em.htg\_pln.\*  
35: em.htg\_rod.\*  
36: em.htg\_mam.\*  
37: em.htg\_vrt.\*  
38: em.sy.\*  
39: em.htgo\_hum.\*  
40: em.htgo\_mus.\*  
41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	86	100.0	549	6	ARI24884	Sequence
2	86	100.0	669	10	AF193843	Mus muscu
3	86	100.0	678	10	MNRNATMS1	X86374 M.musculus
4	86	100.0	680	10	AF076482	Mus muscu
5	86	100.0	713	10	BC005582	Mus muscu
6	81	94.2	630	10	AF154114	Rattus no
7	81	94.2	182897	2	AC110846	AC110846 Rattus no
8	73	84.9	700	4	CDR131676	Camelus d
9	73	84.9	5358	4	CDR409286	Camelus d
10	72	83.7	688	4	AY083309	Bos tauru
11	64	74.4	690	9	AF078483	AF078483 Homo sapi
12	64	74.4	724	9	AF242517	AF242517 Homo sapi
13	64	74.4	166500	9	AC007785	AC007785 Homo sapi
14	53	61.6	46305	9	AC002319	AC002319 Homo sapi
15	53	61.6	87903	9	AL732364	Human DNA
16	53	61.6	271203	9	AL772161	Human DNA
17	52	60.5	193126	9	AL139275	Homo sapi
18	50	58.1	66206	10	AC093315	Human DNA
19	50	58.1	155645	2	AC094964	AC094964 Mus muscu
20	49.5	57.6	176698	2	AC114440	Rattus no
21	49	57.0	59605	9	AC108221	AC108221 Rattus no
22	49	57.0	84633	9	AC022093	Homo sapi
23	49	57.0	115243	9	AL157771	Human DNA
24	49	57.0	122102	9	AC010468	Homo sapi
25	49	57.0	129575	9	AC114945	AC114945 Homo sapi
26	49	57.0	151603	9	AL157361	Human DNA
27	49	57.0	159930	2	AC014935	AC014935 Drosophil
28	49	57.0	161452	2	AC020925	Homo sapi
29	49	57.0	162242	2	AC090803	Homo sapi
30	49	57.0	189288	9	AL583853	Human DNA
31	49	57.0	186002	3	AC006933	AC006933 Drosophil
32	49	57.0	188630	2	AC012606	Homo sapi
33	49	57.0	201313	3	AC010688	AC010688 Drosophil
34	49	57.0	280887	3	AE003526	Drosophil
35	48.5	56.4	164243	2	AC111855	AC111855 Rattus no
36	48.5	56.4	172456	2	AC098515	Rattus no
37	48.5	56.4	249167	2	AC093960	AC093960 Rattus no
38	48	55.8	621	6	AX123016	AX123016 Sequence
39	48	55.8	670	3	AF076481	AF076481 Trichoplu
40	48	55.8	744	6	AX065825	AX065825 Sequence
41	48	55.8	77402	9	HS796111	AL031257 Human DNA
42	48	55.8	110000	2	AC110833_1	Continuation (2 of
43	48	55.8	119958	2	AC104706	AC104706 Oryza sat
44	48	55.8	135424	2	AC069040	AC069040 Homo sapi
45	48	55.8	139931	9	AC078818	AC078818 Homo sapi

ALIGNMENTS

RESULT 1

```

AR124884
LOCUS       AR124884               549 bp    DNA          linear      PAT 16-MAY-2001
DEFINITION   Sequence 1 from patent US 6172211.
ACCESSION   AR124884
VERSION     AR124884.1  GI:14110245
KEYWORDS    Unknown.
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1 (bases 1 to 549)
AUTHORS     Georgiev,G.P., Kiselev,S.L., Prokhorchouk,E.B. and Ostermann,E.
TITLE       Nucleic acid encoding tag7 polypeptide
JOURNAL     Patent: US 6172211-A 1 09-JAN-2001;
FEATURES    Location/Qualifiers
             source
             1..549
             /organism="unknown"
BASE COUNT   117 a   161 c   151 g   120 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2610
2611
2612
2613
261
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 4, 2002, 18:59:21 : Search time 2209 Seconds  
(without alignments)  
117.305 Million cell updates/sec

Title: US-09-462-625-2\_COPY\_145\_160  
Perfect score: 86  
Sequence: 1 RGLRSLNYKVGHRDV 16

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2.1/USPTO\_spool/US09462625/tunat\_04112002\_110248\_23136/app\_query.fasta\_1.199  
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462625 -CGN\_1\_1716 -runat\_04112002\_110248\_23136 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_fod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	86	100.0	274	13	BI159353
2	86	100.0	277	9	AV085455
3	86	100.0	281	9	AA691322
4	86	100.0	305	9	AA930493
5	86	100.0	316	9	AA930504
6	86	100.0	322	9	AI849253
7	86	100.0	329	9	AA616255
8	86	100.0	384	9	AA689633
9	86	100.0	528	9	AA734993
10	86	100.0	534	9	AA734805
11	86	100.0	547	9	AA689893
12	86	100.0	548	9	AI585767
13	86	100.0	548	9	AA597240
14	86	100.0	571	13	BG963242
15	86	100.0	575	9	AI507116
16	86	100.0	580	13	BI556032
17	86	100.0	589	12	BG174272
18	86	100.0	591	9	AA238564
19	86	100.0	600	12	BG871384
20	86	100.0	610	13	BI154844
21	86	100.0	614	10	BE199698
22	86	100.0	619	12	BF302505
23	86	100.0	637	13	BI149595
24	86	100.0	650	13	BI155774
25	86	100.0	676	13	BI650838
26	86	100.0	682	11	AK008335
27	86	100.0	686	14	BQ951956
28	86	100.0	688	12	BG244455
29	86	100.0	703	13	BI453419
30	86	100.0	706	12	BF163190
31	86	100.0	842	13	BI409815
32	83	96.5	284	9	AV092014
33	83	96.5	645	13	BG975104
34	82	95.3	601	9	AA238752
35	81	94.2	226	9	AA082572
36	81	94.2	472	13	BI291182
37	81	94.2	504	12	BF548605
38	81	94.2	506	9	AA875213
39	78	90.7	234	9	AI153056
40	74	86.0	627	9	AA228200
41	72	83.7	555	12	BF076851
42	71	82.6	513	14	BQ564677
43	70	81.4	384	10	BE654595
44	69	80.2	223	9	AV065626
45	67	77.9	224	9	AV062378

# ALIGNMENTS

RESULT 1  
BI159353  
LOCUS 602919837F1 NCI\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5060072 5',  
DEFINITION mRNA sequence.  
ACCESSION BI159353  
VERSION BI159353.1 GI:14619354  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
TITLE NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)





Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:619496  
 Putative full length read  
 vector to vector length is  
 Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence stop: 203.

## FEATURES

source

Location/Qualifiers  
 1. .281  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1138224"  
 /clone\_lib="Barstead mouse irradiated colon MPLRB7"  
 /dev\_stage="8 weeks"  
 /lab\_host="DH10B"  
 /note="vector: p7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: EcoRI; Site\_2: NotI; Tissue obtained  
 from 8 week old mouse. Colon was harvested 72 hours after  
 irradiation with 1400 Gys. 1st strand cDNA was primed  
 with a Not I - oligo(dT) primer  
 [5'TGTACGAATCTGAAGTGGAGCGCCCTTTTTTTTTTTTTTTTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors [AATCGATCCTTG], digested with Not I and cloned  
 into the Not I and Eco RI sites of the modified p7T3  
 vector. Library constructed by Bob Barstead."  
 BASE COUNT 69 a 87 c 67 g 58 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3.89e-07 Length: 281  
 Score: 86.00 Matches: 16  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x AA691322 (1-281)

QY 1 ArgGlyPheLeuArgSerAsnTyrGluVallysglyHisArgAspVal 16  
 |||||  
 Db 73 CGGGCTCTTGAGATCCAACTATGAGTCAAGACACCGGGATGTG 120

## RESULT 4

AA930493  
 LOCUS  
 DEFINITION  
 vo45c06.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA  
 clone IMAGE:1052842 5' similar to gb:X86374 M.musculus mRNA for  
 TAG7 protein (MOUSE);, mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

AA930493 305 bp mRNA linear EST 23-APR-1998  
 house mouse.  
 Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 305)  
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.

TITLE  
JOURNAL  
COMMENT

The WashU-HHMI Mouse EST Project  
 Unpublished (1996)  
 Contact: Marra M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:584418  
 Putative full length read  
 vector to vector length is 306  
 Seq primer: -28m13 rev2 ET from Amersham.

## FEATURES

source

Location/Qualifiers  
 1. .305  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1052842"  
 /clone\_lib="Barstead mouse irradiated colon MPLRB7"  
 /dev\_stage="8 weeks"  
 /lab\_host="DH10B"

/note="vector: p7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: EcoRI; Site\_2: NotI; Tissue obtained  
 from 8 week old mouse. Colon was harvested 72 hours after  
 irradiation with 1400 Gys. 1st strand cDNA was primed  
 with a Not I - oligo(dT) primer  
 [5'TGTACGAATCTGAAGTGGAGCGCCCTTTTTTTTTTTTTTTTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors [AATCGATCCTTG], digested with Not I and cloned  
 into the Not I and Eco RI sites of the modified p7T3  
 vector. Library constructed by Bob Barstead."  
 BASE COUNT 74 a 97 c 69 g 65 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 4.42e-07 Length: 305  
 Score: 86.00 Matches: 16  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x AA930493 (1-305)

QY 1 ArgGlyPheLeuArgSerAsnTyrGluVallysglyHisArgAspVal 16  
 |||||  
 Db 96 CGGGCTCTTGAGATCCAACTATGAGTCAAGACACCGGGATGTG 143

## RESULT 5

AA930504  
 LOCUS  
 DEFINITION  
 vo45d09.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA  
 clone IMAGE:1052849 5' similar to gb:X86374 M.musculus mRNA for  
 TAG7 protein (MOUSE);, mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE

AA930504 316 bp mRNA linear EST 23-APR-1998  
 house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 316)  
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.

## REFERENCE

## AUTHORS

## JOURNAL

## COMMENT

The WashU-HHMI Mouse EST Project  
 Unpublished (1996)  
 Contact: Marra M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:584425

Seq primer: -28ml3 rev2 ET from Amersham.

```

FEATURES             source
  Location/Qualifiers
    1. .316
      /organism="Mus musculus"
      /strain="FVB/N"
      /db_xref="taxon:10090"
      /clone="IMAGE:1052849"
      /clone_lib="Barstead mouse irradiated colon MPLRB7"
      /dev_stage="8 weeks"
      /lab_host="DH10B"
      /note="vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: EcoRI; Site 2: NotI; Tissue obtained
from 8 week old mouse. Colon was harvested 72 hours after
irradiation with 1400 Gys. 1st strand cDNA was primed
with a Not I - oligo(dT) primer
[5'TGTACCAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors [AATTCGATCCTTG], digested with Not I and cloned
into the Not I and Eco RI sites of the modified p7T3
vector. Library constructed by Bob Barstead."
BASE COUNT          96 a 90 c 70 g 60 t
ORIGIN
Alignment Scores:
Pred. No.:          4.67e-07          Length:          316
Score:              86.00             Matches:          16
Percent Similarity: 100.00%           Conservative:    0
Best Local Similarity: 100.00%         Mismatches:     0
Query Match:        100.00%           Indels:          0
DB:                  9                Gaps:            0

US-09-462-625-2_COPY_145_160 (1-16) x AA930504 (1-316)

Qy 1 ArgGlyPheLeuArgSerAsnTyrGluVallyGlyHisArgAspVal 16
|||||
Db 78 CGGGCTTCTTGAGATCCAACTATGACGTCAAGGACACCGGGGATGTG 125

RESULT 6
AI849253/c
LOCUS          AI849253          322 bp      mRNA          linear          EST 15-JUL-1999
DEFINITION    UI-M-AJ1-ag2-e-10-0-UI.s1 NIH_BMAP_MOB_N Mus musculus cDNA clone
               UI-M-AJ1-ag2-e-10-0-UI 3', mRNA sequence.
ACCESSION     AI849253
VERSION       AI849253.1 GI:5493159
KEYWORDS      EST.
SOURCE        house mouse.
ORGANISM      Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 322)
AUTHORS       Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE         Normalization and subtraction: two approaches to facilitate gene
               discovery
JOURNAL       Genome Res. 6 (9), 791-806 (1996)
MEDLINE       97044477
COMMENT       Contact: Chhin, H
               National Institute of Mental Health
               6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
               20892-9643, USA
               Tel: 301 443 1706
               Fax: 301 443 9890
               Email: mEST@mail.nih.gov
               The sequence contained an oligo-dT track that was present in the
               oligonucleotide that was used to prime the synthesis of first
               strand cDNA and therefore this may represent a bonafide poly A
               tail. The sequence tag present in the cDNA between the NotI site
               and the oligo-dT track served to verify it as a clone from the
               normalized olfactory bulbs library cDNA Library Preparation: M.B.
               Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
               available by the means that is soon to be determined. When NIH
               determines the means for distribution of the BMAP cDNA clones, this
               record will be updated accordingly when that means is determined.
               Seq primer: M13 Forward

```

```

FEATURES             source
  Location/Qualifiers
    1. .322
      /organism="Mus musculus"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="UI-M-AJ1-ag2-e-10-0-UI"
      /clone_lib="NIH_BMAP_MOB_N"
      /dev_stage="27-32 days"
      /lab_host="DH10B (Life Technologies)"
      /note="vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_MOB_N library is a normalized library constructed
from mouse olfactory bulbs. The tag is a string of 5
nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described
by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
, 1996. Tissue provided by Ms. Annie Novakovich,
Zivic-Miller Laboratories.
TAG LIB=NIH_BMAP_MOB_N
TAG_TISSUE=olfactory-bulbs
TAG_SEQ=CATGG"
BASE COUNT          62 a 70 c 96 g 94 t
ORIGIN
Alignment Scores:
Pred. No.:          4.81e-07          Length:          322
Score:              86.00             Matches:          16
Percent Similarity: 100.00%           Conservative:    0
Best Local Similarity: 100.00%         Mismatches:     0
Query Match:        100.00%           Indels:          0
DB:                  9                Gaps:            0

US-09-462-625-2_COPY_145_160 (1-16) x AI849253 (1-322)

Qy 1 ArgGlyPheLeuArgSerAsnTyrGluVallyGlyHisArgAspVal 16
|||||
Db 230 CGGGCTTCTTGAGATCCAACTATGACGTCAAGGACACCGGGGATGTG 183

RESULT 7
AA616255
LOCUS          AA616255          329 bp      mRNA          linear          EST 07-OCT-1997
DEFINITION    vo50e07.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA
               clone IMAGE:1053348 5' similar to gb:X86374 M.musculus mRNA for
               TAG7 protein (MOUSE);, mRNA sequence.
ACCESSION     AA616255
VERSION       AA616255.1 GI:2503460
KEYWORDS      EST.
SOURCE        house mouse.
ORGANISM      Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 329)
AUTHORS       Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
               Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
               Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
               Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
               Waterston,R.
TITLE         The WashU-HHMI Mouse EST Project
JOURNAL       Unpublished (1996)
COMMENT       Contact: Marra M/Mouse EST Project
               WashU-HHMI Mouse EST Project
               Washington University School of Medicine#
               4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
               Tel: 314 286 1800
               Fax: 314 286 1810
               Email: mouseest@watson.wustl.edu
               This clone is available royalty-free through LNL ; contact the
               IMAGE Consortium (info@image.llnl.gov) for further information.
               MGI:584924
               Putative full length read
               vector to vector length is
               Seq primer: -28ml3 rev2 ET from Amersham

```

High quality sequence stop: 227.

## FEATURES

## source

Location/Qualifiers  
1. .329  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1053348"  
/clone\_lib="Barstead mouse irradiated colon MPLRB7"  
/dev\_stage="8 weeks"  
/lab\_host="DH10B"

/note="vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; Tissue obtained from 8 week old mouse. Colon was harvested 72 hours after irradiation with 1400 Gys. 1st strand cDNA was primed with a Not I - oligo(dT) primer  
[5'TGTACGAATCTGAAGTGGAGCGCCGCTCTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATCGGATCCTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library constructed by Bob Barstead."

BASE COUNT 77 a 100 c 79 g 72 t 1 others

## ORIGIN

Alignment Scores:  
Pred. No.: 4.97e-07 Length: 329  
Score: 86.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x AA616255 (1-329)

Qy 1 ArgGlyPheLeuArgSerAsnTyrGluValVallysGlyHisArgAspVal 16

Db 143 CGGGCTCTTCTGAGATCAACTATGATCAAGTCAAGGACACCGGGATGTG 190

## RESULT 8

## AA689633

## LOCUS

DEFINITION vs09g08.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA  
clone IMAGE:1137758 5' similar to gb:X86374 M.musculus mRNA for TAG7 protein (MOUSE);, mRNA sequence.

AA689633 384 bp mRNA EST 16-DEC-1997

VERSION clone IMAGE:1137758 5' similar to gb:X86374 M.musculus cDNA

KEYWORDS TAG7 protein (MOUSE);, mRNA sequence.

SOURCE EST.

ORGANISM house mouse.

REFERENCE Mus musculus

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 384)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

The WashU-HHMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:619030

Seq primer: -28ml3 rev2 ET from Amersham.

Location/Qualifiers

1. .384

/organism="Mus musculus"

/strain="FVB/N"

## FEATURES

## source

Location/Qualifiers  
1. .384  
/organism="Mus musculus"  
/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:1137758"

/clone\_lib="Barstead mouse irradiated colon MPLRB7"

/dev\_stage="8 weeks"

/lab\_host="DH10B"

/note="vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; Tissue obtained from 8 week old mouse. Colon was harvested 72 hours after irradiation with 1400 Gys. 1st strand cDNA was primed with a Not I - oligo(dT) primer

[5'TGTACGAATCTGAAGTGGAGCGCCGCTCTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATCGGATCCTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library constructed by Bob Barstead."

BASE COUNT 96 a 102 c 102 g 84 t

## ORIGIN

Alignment Scores:  
Pred. No.: 6.32e-07 Length: 384  
Score: 86.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x AA689633 (1-384)

Qy 1 ArgGlyPheLeuArgSerAsnTyrGluValVallysGlyHisArgAspVal 16

Db 269 CGGGCTCTTCTGAGATCAACTATGATCAAGTCAAGGACACCGGGATGTG 316

## RESULT 9

## AA734993

## LOCUS

DEFINITION vs17h04.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA  
clone IMAGE:1138519 5' similar to gb:X86374 M.musculus mRNA for TAG7 protein (MOUSE);, mRNA sequence.

AA734993 528 bp mRNA linear EST 07-JAN-1998

VERSION clone IMAGE:1138519 5' similar to gb:X86374 M.musculus cDNA

KEYWORDS TAG7 protein (MOUSE);, mRNA sequence.

SOURCE EST.

ORGANISM house mouse.

REFERENCE Mus musculus

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 528)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

The WashU-HHMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:619791

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 467.

Location/Qualifiers

1. .528

/organism="Mus musculus"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:1138519"

/clone\_lib="Barstead mouse irradiated colon MPLRB7"

/dev\_stage="8 weeks"

irradiation with 1400 Gys. 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'TGTTAGTCTCAAGTACGAGGCGCCGCTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATTCGATCCTTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead."

BASE COUNT	130 a	156 c	137 g	111 t
ORIGIN				
Alignment Scores:				
pred. No.:	1.06e-06			534

```

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-462-625-2_COPY_145_160 (1-16) x AA734805 (1-534)

Qy . 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16
Db 326 CGGGGCTTCTTGATGCCACTATGTAAGTCAAGAGCACCGGGATGTG 373

```

RESULT	11
AA689693	
LOCUS	547 bp mRNA linear EST 16-DEC-1997
DEFINITION	v503409.1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA clone IMAGE1137137 5' similar to gb:X86374 M.musculus mRNA for TAG7 protein (MOUSE);,, mRNA sequence.
ACCSSION	AA689693
VERSION	AA689693.1 GI:2690629
KEYWORDS	EST.
SOURCE	house mouse.

REFERENCE	<p>1 (bases 1 to 547)</p> <p>Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Weising, B., Wylie, T., Lennon, G., Soares, B., Willson, R. and Waterston, R.</p>
TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	<p>Contact: Maria M/Mouse EST Project</p> <p>WashU-HHMI Mouse EST Project</p> <p>Washington University School of Medicine</p> <p>4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108</p> <p>Tel: 314 286 1800</p>

fax: 314 200 1010  
 Email: mouse@wustl.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:618409  
 Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 141.  
 Location/Qualifiers  
 source  
 1..547

```

/strain="v5b/a"
/db_xref="taxon:10090"
/Clone="IMAGE:1137137"
/Clone_lib="Barstead mouse irradiated colon MPLRB7"
/dev_stage="8 weeks"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
from 8 week old mouse. Colon was harvested 72 hours after
irradiation with 1400 Gys. First strand cDNA was primed
with a Not I - oligo(dT) primer
[5'TGTACGAATCTCAAGTGGAGGGCGGCGCTTTTTTTTTTTTTTTTTT
T 3'; double-stranded cDNA was ligated to Eco RI

```

adaptors [AATTCGATCCTTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed by Bob Barstead "

BASE COUNT	129 a	159 c	138 g	121 t
ORIGIN				

Alignment Scores:	
Pred. No.:	1.1e-06
Score:	86.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	9
Length:	547
Matches:	16
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-09-462-625-2\_COPY\_145\_160 (1-16) x AA689693 (1-547)

Qy 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16  
 |||||  
 Db 352 CGGSGCTCTTGAGATCCAACATATGAAGTCAAAGSCAACCCGGGATGTG 399

RESULT 12  
AI585767/c

LOCUS	548 bp	linear	EST 06-APR-1999
DEFINITION	Barstead mouse irradiated colon MPERB7 Mus musculus cDNA clone IMAGE:1137137 3' similar to gb:X86374 M.musculus mRNA for TAG7 protein (MOUSE); mRNA sequence		

ACCESSION	AI585767
VERSION	AI585767.1
KEYWORDS	EST.

KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 548)

REFERENCE

AUTHORS

Marrs, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person  
B., Swaller, T., Gibson, M., Page, D., Harvey, N., Schurk, R., Ritter  
E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,

TITLE	COMMENT
The WashU-NCI Mouse EST Project 1999	Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999	Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis MO 63108	USA

Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:618409  
This clone was previously sequenced on the 5' end only, this new  
data is from the 3' end  
High quality sequence span: 260

```

FEATURES
source
1. .548
Location/Qualifiers
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="1137137"
/clone_lib="Barstead mouse irradiated colon MPLRB7"
/dev_stage="8 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI. Tissue obtained
from 8 week old mouse. Colon was harvested 72 hours after
irradiation with 1400 Gys. 1st strand cDNA was primed
with a Not I - oligo(dT) primer
[5'TGTTCAAGTCTGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors [AATTCGATCCTTGG], digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library constructed by Bob Barstead."
110 a 137 c 163 g 137 t 1 others
BASE COUNT

```

BASE COUNT	110 a	137 c	163 g	137 t	1 others
BASE COUNT	110 a	137 c	163 g	137 t	1 others

## Alignment Scores:

Pred. No.: 1.1e-06 Length: 548  
 Score: 86.00 Matches: 16  
 Percent Similarity: 100.00% Conservativeness: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0

US-09-462-625-2\_COPY\_145\_160 (1-16) x AA597240 (1-548)

Qy 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16  
 |||||||  
 Db 342 CGGGCTTCTTGAGATCCAACTATGAGTCAAGAGCACCGGGATGTG 389

## RESULT 14

LOCUS BG963242 571 bp mRNA linear EST 12-JUN-2001  
 DEFINITION 602828323F1 NCI\_CGAP\_Co24 Mus musculus cDNA clone IMAGE:4983340 5', mRNA sequence.

ACCESSION BG963242

VERSION BG963242.1 GI:14350892

## KEYWORDS

EST.

SOURCE house mouse.

## ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 571)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Tissue Procurement: Jeffrey E. Green, M.D.

CDNA Library Preparation: Life Technologies, Inc.

CNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LHAM10987 row: m column: 05

High quality sequence stop: 312.

Location/Qualifiers

1..571

/organism="Mus musculus"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:4983340"

/lab\_host="NCI\_CGAP\_Co24"

/note="Organ: colon; Vector: pCMV-SPORT6; Site.1: NotI;

Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.6 kb. Constructed by Life

Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 199 a 203 c 111 g 58 t

ORIGIN

Alignment Scores:

Pred. No.: 1.17e-06 Length: 571  
 Score: 86.00 Matches: 16  
 Percent Similarity: 100.00% Conservativeness: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 13

US-09-462-625-2\_COPY\_145\_160 (1-16) x BG963242 (1-571)

Qy 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16  
 |||||||  
 Db 53 CGGGCTTCTTGAGATCCAACTATGAGTCAAGAGCACCGGGATGTG 100

## RESULT 15

LOCUS AI507116/c 575 bp mRNA linear EST 11-MAR-1999

AI507116

## DEFINITION

Vj82f09.x1 Soares\_mammary\_gland\_NbMMG Mus musculus cDNA clone  
 IMAGE:935561 3' similar to gb:X86374 M.musculus mRNA for TAG7  
 protein (MOUSE);, mRNA sequence.

## ACCESSION

AI507116

## VERSION

AI507116.1

## KEYWORDS

EST.

## SOURCE

house mouse.

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 575)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:540481

This clone was previously sequenced on the 5' end only, this new

data is from the 3' end

High quality sequence stop: 349.

Location/Qualifiers

1..575

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:935561"

/clone\_lib="Soares\_mammary\_gland\_NbMMG"

/sex="male"

/tissue\_type="mammary gland"

/dev\_stage="4 weeks"

/lab\_host="DH10B"

/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia

RI; with a modified polylinker; Site.1: Not I; Site.2: Eco

RI; 1st strand cDNA was primed with a Not I - Oligo(dt)

primer [5'

TGTTCACCAATCTGAAGTGGGAGCGCGCGAATGTTTGTGTGTGTGTGTGTGTGT

T 3']; Double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pT7T3 vector.

RNA provided by Dr. Minoru Ko, Wayne State Univ. Library

constructed and normalized by Bento Soares and M.Fatima

Bonaldo."

BASE COUNT 114 a 146 c 176 g 138 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 1.18e-06 Length: 575  
 Score: 86.00 Matches: 16  
 Percent Similarity: 100.00% Conservativeness: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 9

US-09-462-625-2\_COPY\_145\_160 (1-16) x AI507116 (1-575)

Qy 1 ArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAspVal 16  
 |||||||  
 Db 219 CGGGCTTCTTGAGATCCAACTATGAGTCAAGAGCACCGGGATGTG 172

Search completed: November 4, 2002, 20:31:04  
 Job time : 2214 secs